(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) **EP 4 109 058 A1**

(12) **EUROPEAN PATENT APPLICATION**
published in accordance with Art. 153(4) EPC

(43) Date of publication:
28.12.2022 Bulletin 2022/52

(21) Application number: **20920020.3**

(22) Date of filing: **20.02.2020**

(51) International Patent Classification (IPC):
G01H 3/08 *(2006.01)* ***G01M 99/00*** *(2011.01)*
G06T 7/00 *(2017.01)* ***G10L 25/30*** *(2013.01)*
G10L 25/51 *(2013.01)*

(52) Cooperative Patent Classification (CPC):
G01H 3/08; G01M 99/00; G06T 7/00; G10L 25/30; G10L 25/51

(86) International application number:
PCT/IB2020/000097

(87) International publication number:
WO 2021/165712 (26.08.2021 Gazette 2021/34)

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR
Designated Extension States:
BA ME
Designated Validation States:
KH MA MD TN

(71) Applicants:
- **NISSAN MOTOR Co., Ltd.**
  Yokohama-shi, Kanagawa 221-0023 (JP)
- **Renault s.a.s**
  92100 Boulogne Billancourt (FR)

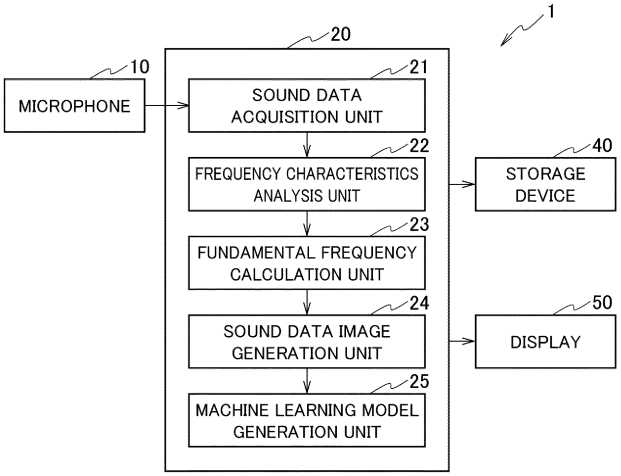
(72) Inventors:
- **HIROSE, Satoru**
  Atsugi-shi, Kanagawa 243-0123 (JP)
- **TAKAGI, Toru**
  Atsugi-shi, Kanagawa 243-0123 (JP)

(74) Representative: **Osha Liang**
2, rue de la Paix
75002 Paris (FR)

(54) IMAGE PROCESSING DEVICE AND IMAGE PROCESSING METHOD

(57) An image processing apparatus (1) includes a controller (20). The controller (20) calculates a fundamental frequency component included in sound data and a harmonic component corresponding to the fundamental frequency component, converts the fundamental frequency component and the harmonic component into image data, and generates a sound image where the fundamental frequency component and the harmonic component converted into the image data are arranged adjacent each other.

FIG. 1



EP 4 109 058 A1

Description

TECHNICAL FIELD

[0001] The present invention relates to an image processing apparatus and an image processing method.

BACKGROUND ART

[0002] A method for determining whether a sound of a device is normal or abnormal has been known (Patent Literature 1). The invention described in Patent Literature 1 determines whether a sound of a device is normal or abnormal using a locus vector indicating intensity features in all time directions and a previously learned identification parameter.

CITATION LIST

PATENT LITERATURE

[0003] Patent Literature 1: WO 2015/068446

SUMMARY OF INVENTION

TECHNICAL PROBLEM

[0004] When a machine learning model for determining an abnormality using sound data is generated, it is conceivable to generate a machine learning model by quantifying sound data by physical quantities, such as sound pressure (dB) indicating the magnitude of sound data and frequency (Hz) indicating the pitch of sound data. Although it is possible to effectively generate a machine learning model by imaging sound data, the invention described in Patent Literature 1 does not mention imaging of sound data.

[0005] In response to the above issue, an object of the present invention is to provide an image processing apparatus and an image processing method of imaging sound data.

TECHNICAL SOLUTION

[0006] An image processing apparatus according to one aspect of the present invention calculates a fundamental frequency component included in sound data and a harmonic component corresponding to the fundamental frequency component, converts the fundamental frequency component and the harmonic component into image data, and generates a sound image where the fundamental frequency component and the harmonic component that have been converted into the image data are arranged adjacent each other.

ADVANTAGEOUS EFFECTS

[0007] The present invention enables a machine learning model of sound data to be generated by imaging sound data.

BRIEF DESCRIPTION OF DRAWINGS

[0008]

[Fig. 1] Fig. 1 is a schematic configuration diagram illustrating an image processing apparatus 1 according to a first embodiment of the present invention.
[Fig. 2A] Fig. 2A is an example of an image of sound data according to the first embodiment of the present invention.
[Fig. 2B] Fig. 2B is an example of a frequency analysis result.
[Fig. 3] Fig. 3 is an example of a machine learning model displayed on a display 50.
[Fig. 4] Fig. 4 is a flowchart illustrating an operation example of the image processing apparatus 1 according to the first embodiment of the present invention.
[Fig. 5] Fig. 5 is a schematic configuration diagram illustrating an image processing apparatus 1 according to a second embodiment of the present invention.
[Fig. 6] Fig. 6 is an example of an image of sound data according to the second embodiment of the present invention.
[Fig. 7] Fig. 7 is a flowchart illustrating an operation example of the image processing apparatus 1 according to the second embodiment of the present invention.
[Fig. 8] Fig. 8 is a schematic configuration diagram illustrating an image processing apparatus 1 according to a third embodiment of the present invention.
[Fig. 9] Fig. 9 is an example of an image of sound data according to the third embodiment of the present invention.
[Fig. 10] Fig. 10 is a flowchart illustrating an operation example of the image processing apparatus 1 according to the third embodiment of the present invention.
[Fig. 11] Fig. 11 is a schematic configuration diagram illustrating an image processing apparatus 1 according to a fourth embodiment of the present invention.
[Fig. 12] Fig. 12 is an example of an image of sound data according to the fourth embodiment of the present invention.
[Fig. 13] Fig. 13 is a flowchart illustrating an operation example of the image processing apparatus 1 according to the fourth embodiment of the present invention.
[Fig. 14] Fig. 14 is a schematic configuration diagram of an image processing apparatus 1 according to a fifth embodiment of the present invention.
[Fig. 15] Fig. 15 is a diagram illustrating an example of a method of cutting out an image of sound data.
[Fig. 16] Fig. 16 is a flowchart illustrating an operation example of the image processing apparatus 1 ac-

cording to the fifth embodiment of the present invention.

[Fig. 17] Fig. 17 is a schematic configuration diagram of an image processing apparatus 1 according to a sixth embodiment of the present invention.

[Fig. 18A] Fig. 18A is a flowchart illustrating an operation example of the image processing apparatus 1 according to the sixth embodiment of the present invention.

[Fig. 18B] Fig. 18B is a flowchart illustrating an operation example of the image processing apparatus 1 according to the sixth embodiment of the present invention.

DESCRIPTION OF EMBODIMENTS

[0009] Embodiments of the present invention are described below with reference to the drawings. In the drawings, the same parts are denoted by the same reference numerals, and the description thereof is omitted.

[First embodiment]

(Configuration example of image processing apparatus)

[0010] A configuration example of an image processing apparatus 1 according to a first embodiment is described with reference to Fig. 1. As illustrated in Fig. 1, the image processing apparatus 1 includes a controller 20, a storage device 40, and a display 50.

[0011] The controller 20 acquires sound data via a microphone 10 and analyzes the acquired sound data. The controller 20 is a general-purpose microcomputer including a CPU (central processing unit), a memory, and an input/output unit. The microcomputer has a computer program installed to function as the image processing apparatus 1. By executing the computer program, the microcomputer functions as multiple information processing circuits provided in the image processing apparatus 1. Note that the example described here is to use software to realize the multiple information processing circuits provided in the image processing apparatus 1; however, it is also possible to prepare dedicated hardware for executing each information processing described below to configure the information processing circuits. Further, the multiple information processing circuits may be configured by individual hardware. The controller 20 includes a sound data acquisition unit 21, a frequency characteristics analysis unit 22, a fundamental frequency calculation unit 23, a sound data image generation unit 24, and a machine learning model generation unit 25, as the multiple information processing circuits.

[0012] As described above, the sound data acquisition unit 21 acquires sound data via the microphone 10. The sound data acquired by the sound data acquisition unit 21 are converted into an electric signal and treated as time-series data. The sound data acquired by the sound data acquisition unit 21 is given a label indicating normal or abnormal and is used as machine learning teacher data. Note that the sound data is, for example, sound data of a machine used in a factory.

[0013] The frequency characteristics analysis unit 22 analyzes frequency characteristics of the sound data acquired by the sound data acquisition unit 21. As an analysis method, signal processing represented by FFT (Fast Fourier Transform) is used, for example. The FFT converts time series data into frequency data to provide "frequency-energy characteristics".

[0014] The fundamental frequency calculation unit 23 calculates a fundamental frequency of the sound data using the "frequency-energy characteristics" acquired by the frequency characteristics analysis unit 22. The fundamental frequency calculation unit 23 calculates harmonics corresponding to the calculated fundamental frequency.

[0015] The sound data image generation unit 24 generates an image of the sound data using the fundamental frequency calculated by the fundamental frequency calculation unit 23.

[0016] The machine learning model generation unit 25 generates an optimum machine learning model for determining normality and abnormality in the sound data using the image generated by the sound data image generation unit 24. The machine learning model generation unit 25 confirms the performance of determining normality and abnormality of sound data by multiple machine learning algorithms using cross verification or the like and outputs a model with the best performance. Note that the machine learning model generation unit 25 may also output other models together.

[0017] The storage device 40 stores the "frequency-energy characteristics" acquired by the frequency characteristics analysis unit 22, the sound data image generated by the sound data image generation unit 24, the machine learning model generated by the machine learning model generation unit 25, and the like.

[0018] The display 50 displays the machine learning model generated by the machine learning model generation unit 25 and its prediction performance and the like.

[0019] Next, an example of the image generated by the sound data image generation unit 24 is described with reference to Figs. 2A and 2B.

[0020] An image 60 in Fig. 2A is an example of the image generated by the sound data image generation unit 24. First, a method of generating the image 60 is described. Details of the image 60 is described later. The sound data acquisition unit 21 acquires sound data of any machine. The sound data is converted into an electric signal. Note that the image 60 is acquired by converting sound data into image data and may be referred to as a sound image.

[0021] The frequency characteristics analysis unit 22 analyzes the electric signal using the FFT. Fig. 2B illustrates an example of an analysis result. In Fig. 2B, the vertical axis indicates sound pressures (dB(A)), and the horizontal axis indicates frequencies (Hz). The FFT pro-

vides the "frequency-energy characteristics" illustrated in Fig. 2B.

[0022] The fundamental frequency calculation unit 23 calculates the fundamental frequency of sound data using the "frequency-energy characteristics" illustrated in Fig. 2B. The fundamental frequency refers to a frequency of the lowest frequency component when the signal (which refers to sound data converted into an electric signal here) is represented by the synthesis of sine waves. The fundamental frequency is sometimes called a fundamental wave. According to the FFT, as illustrated in Fig. 2B, multiple peak frequencies (peaks of a spectrum) generally appears. The method of calculating the fundamental frequency is not limited, but for example, as illustrated in Fig. 2B, the minimum peak frequency (400Hz in Fig. 2B) among the multiple peak frequencies may be calculated as the fundamental frequency.

[0023] As another calculation method, a frequency between peaks may be calculated as the fundamental frequency. As illustrated in Fig. 2B, large peaks appear at 1200 Hz and 1600 Hz. The frequency between the peaks (1600 - 1200 = 400 Hz) may be calculated as the fundamental frequency.

[0024] As another calculation method, a difference of frequencies between peaks may be acquired, and when the difference is the same as the minimum peak frequency, the frequency may be calculated as the fundamental frequency. Specifically, as illustrated in Fig. 2B, since the difference in frequency between the peaks (400 Hz as described above) is the same as the minimum peak frequency (400 Hz), the fundamental frequency calculation unit 23 may calculate 400 Hz as the fundamental frequency. Here, "the difference is the same as the minimum peak frequency" is not limited to the exact same. "The difference is the same as the minimum peak frequency" means substantially the same (can be regarded as nearly the same). In the first embodiment, the fundamental frequency is set as 400 Hz. The fundamental frequency is one per sound data.

[0025] Next, the fundamental frequency calculation unit 23 calculates harmonics corresponding to the calculated fundamental frequency. Harmonics means a higher-order frequency component that includes integral multiples of a wave motion having a certain frequency component (here, the fundamental frequency component). Harmonics are sometimes called overtones in the field of music. As illustrated in Fig. 2B, large peaks are observed at frequencies that are integral multiples (2x, 3x, etc.) of the fundamental frequency (400 Hz). The fundamental frequency calculation unit 23 calculates frequencies (800 Hz, 1200 Hz, etc.) of integral multiples of as harmonics corresponding to the fundamental frequency.

[0026] The sound data image generation unit 24 generates a two-dimensional image, such as a so-called heat map, as illustrated in Fig. 2A, using the fundamental frequency and harmonics calculated by the fundamental frequency calculation unit 23. In Fig. 2A, the vertical axis represents overtones and the horizontal axis represents frequencies. More specifically, the horizontal axis indicates frequencies of 0 to 400 Hz. The overtones on the vertical axis mean overtones of the fundamental frequency (400 Hz). The image 60 is formed by multiple pixels indicating corresponding frequencies. In the image 60, under a pixel 60a indicating 400 Hz, a pixel 60b indicating 800 Hz (2 times 400 Hz) is arranged adjacently. The same applies below, and under the pixel 60b indicating 800 Hz, a pixel indicating 1200 Hz (3 times 400 Hz) is arranged adjacently. In the lower right of the image 60, a pixel indicating 8400 Hz (21 times 400 Hz) is arranged. In the horizontal axis and the vertical axis, the pixels are arranged in frequency order. The image 60 may be image data made from a two-dimensional matrix in which two-dimensional predetermined areas (areas where the pixels 60a, 60b are arranged) are set for respective frequency components, and these predetermined areas are arranged adjacently in the vertical and horizontal directions.

[0027] In this way, the sound data image generation unit 24 converts the fundamental frequency component and harmonic component calculated by the fundamental frequency calculation unit 23 into image data. The fundamental frequency component converted into image data is represented as the pixel 60a (400 Hz). The harmonic component converted into the image data is represented as the pixel 60b (800 Hz). As illustrated in Fig. 2A, the image 60 (sound image) is generated in which the fundamental frequency component (pixel 60a indicating 400 Hz) and the harmonic component (pixel 60b indicating 800 Hz) converted into image data are arranged adjacently. The sound image is acquired by converting the fundamental frequency component included in sound data and the harmonic component into an image by predetermined processing.

[0028] The image 60 is illustrated in a rectangular shape but is not limited to this. For example, the image 60 may have a thin comb shape for increasing resolution. Each of the pixels is also illustrated in a rectangle shape but is not limited to this. Further, the respective pixels are discretely arranged at a distance from each other but not limited thereto. The respective pixels may be continuously arranged without any distance from each other. The vertical axis and the horizontal axis may be interchanged. In Fig. 2A, the fundamental frequency component and the harmonic component corresponding to the fundamental frequency component are arranged adjacently in the vertical axis but may be arranged adjacently in the horizontal axis.

[0029] In Fig. 2A, each pixel is set as a brightness or color corresponding to the amplitude of the sound data. For example, the pixels 60a and 60b having large sound pressures are set to a dark color (for example, red), such as in a heat map. Other pixels (for example, pixel 60c) are set to a light color (for example, green), such as in a heat map. Note that the pixels 60d and 60e are examples of indication of an abnormality in sound data, and brightnesses or colors are set differently from other pixels. The

brightnesses or colors of pixels are set so that the operator can grasp the intensity, normality, abnormality, and the like of the sound data at a glance.

[0030] Next, a machine learning model is described with reference to Fig. 3. The machine learning model generation unit 25 generates an optimum machine learning model for determining normality and abnormality in sound data using the image 60 generated by the sound data image generation unit 24. As described above, the label indicating normal or abnormal is attached to the sound data. Thus, the machine learning in the first embodiment is so-called supervised learning.

[0031] The machine learning model generation unit 25 generates a machine learning model using a well-known machine learning algorithm. Machine learning algorithms to be used include Decision Tree, Random Forest, Gradient Boosted Tree (GBT), General Linear Regression (GLR), Support Vector Machine (SVM), and Deep Learning. However, the present invention is not limited thereto, and any algorithm capable of supervised learning may be used.

[0032] The machine learning prediction performance is compared by any one of or a combination of items indicating the performance of the machine learning algorithm, such as Accuracy, Precision, and Recall. The model with the highest prediction performance is selected based on the comparison result.

[0033] As illustrated in Fig. 3, the machine learning model generation unit 25 displays the generated machine learning model and its prediction performance on the display 50. On the left side of the display 50, the generated machine learning model (Decision Tree, etc.) and its prediction performance (70%, etc.) are displayed. On the right side of the display 50, details of the model (Deep Learning) with the highest prediction performance are displayed. As the details of Deep Learning, the form, number of learning, ratio of normal and abnormal data, number of intermediate layers, and a Confusion Matrix acquired by cross validation.

[0034] Next, an operation example of the image processing apparatus 1 according to the first embodiment is described with reference to a flowchart of Fig. 4.

[0035] In step S101, the sound data acquisition unit 21 acquires sound data via the microphone 10. The acquired sound data are converted into an electric signal and treated as time series data. The process proceeds to step S103, and the frequency characteristics analysis unit 22 analyzes the electric signal acquired in step S101 using the FFT. The FFT provides the "frequency-energy characteristics" (see Fig. 2B).

[0036] The process proceeds to step S105, and the fundamental frequency calculation unit 23 calculates a fundamental frequency of the sound data using the "frequency-energy characteristics" acquired in step S103. The above-described method is used for calculating the fundamental frequency. The process proceeds to step S107, and the sound data image generation unit 24 generates the image 60, such as a heat map, using the fundamental frequency calculated in step S105 (see Fig. 2A). In the image 60, a fundamental frequency component and a harmonic component corresponding to the fundamental frequency component are arranged adjacently.

[0037] The process proceeds to step S109, and the image 60 generated in step S107 is displayed on the display 50. Each of the pixels forming the image 60 is set as a brightness or color corresponding to an amplitude of the sound data. Thus, the operator who sees the image 60 can grasp the intensity, normality, abnormality, and the like of the sound data at a glance.

[0038] The process proceeds to step Sill, and the machine learning model generation unit 25 selects an algorithm for generating a machine learning model. The selected algorithm includes Decision Tree, Random Forest, Gradient Boosted Tree (GBT), General Liner Regression (GLR), Support Vector Machine (SVM), and Deep Learning.

[0039] The process proceeds to step S113, and the machine learning model generation unit 25 generates a machine learning model using the algorithm selected in step Sill. The process proceeds to step S115, and the machine learning model generation unit 25 displays the generated machine learning model and its prediction performance on the display 50 (see Fig. 3). A series of processing is repeatedly executed until completion (step S117).

(Advantageous effects)

[0040] As described above, the image processing apparatus 1 according to the first embodiment provides the following advantageous effects.

[0041] The fundamental frequency calculation unit 23 calculates a fundamental frequency component included in sound data and a harmonic component corresponding to the fundamental frequency component. The sound data image generation unit 24 converts the fundamental frequency component and harmonic component calculated by the fundamental frequency calculation unit 23 into image data. The sound data image generation unit 24 generates the image 60 (sound image) in which the fundamental frequency component (pixel 60a in Fig. 2A) and harmonic component (pixel 60b in Fig. 2A) converted into image data are arranged adjacently. By generating a sound image having such an arrangement, it is possible to effectively generate a machine learning model. Thus, it is possible to clearly display or process the amount of sensation, such as a timbre, which cannot be digitized by a time-series sound signal or spectral characteristics by FFT.

[0042] The fundamental frequency component (pixel 60a) and the harmonic component (pixel 60b) have an overtone relationship. Monophonic tones, such as those of automobile horns and stringed instruments, are influenced by their overtones. According to the first embodiment, such a relationship can be displayed as a sound

image.

[0043] The sound data image generation unit 24 arranges the fundamental frequency component (pixel 60a) and the harmonic component (pixel 60b) adjacent to each other on the vertical axis or the horizontal axis of the sound image. This enables the relationship between the fundamental frequency and the harmonic component to be displayed as a two-dimensional sound image.

[0044] The sound data image generation unit 24 generates a sound image by converting the fundamental frequency component and the harmonic component into image data where brightnesses or colors corresponding to amplitudes of the sound data are set. This enables the intensity of the sound or the like to be clearly displayed as a sound image.

[0045] The sound data image generation unit 24 arranges multiple frequency components of the sound data in the order of frequency on the vertical axis or the horizontal axis of the sound image. This enables the sound data to be displayed as a two-dimensional sound image.

[Second embodiment]

[0046] Next, a second embodiment of the present invention is described with reference to Fig. 5. The second embodiment differs from the first embodiment in that the controller 20 includes a scale setting unit 26. A description of a configuration that overlaps the first embodiment is omitted with reference to reference numerals. The following description focuses on the difference.

[0047] The scale setting unit 26 uses the "frequency-energy characteristics" acquired by the frequency characteristics analysis unit 22 to set a twelve-tone scale of "C, C#, D, D#, E, F, G, G#, A, A#, B, B#, (C)", which corresponds to the "do, re, mi, fa, sol, la, ti (do)" used in music, and octaves (1 to 10 octaves), which are overtones. The scale setting unit 26 classifies sounds of respective frequencies into the twelve-tone scale using the twelve-tone scale and octaves.

[0048] Next, an example of an image generated by the sound data image generation unit 24 is described with reference to Fig. 6.

[0049] An image 61 in Fig. 6 is an example of an image generated by the sound data image generation unit 24. The sound data image generation unit 24 generates the image 61 by arranging pixels indicating frequencies so as to correspond to the twelve-tone scale using the twelve-tone scale and octaves set by the scale setting unit 26 and the frequencies classified by the scale setting unit 26. In Fig. 6, the horizontal axis indicates the twelve-tone scale, and the vertical axis indicates octaves (overtones of frequencies). The first line of the image 61 indicates a scale from 32 Hz to 65 Hz, and as the line goes down, it indicates overtones. Note that in the example in Fig. 2, overtones have been described as integral multiples of the fundamental frequency, but in the example in Fig. 6, an overtone means that it is twice the frequency of the previous one.

[0050] In the second embodiment, the fundamental frequency is not 400 Hz but any value. A pitch name corresponding to the pixel 60a indicating the fundamental frequency is "A". A pitch name corresponding to the pixel 60b indicating a harmonic corresponding to the fundamental frequency is also "A". As illustrated in Fig. 6, the pixel 60a and the pixel 60b are arranged adjacent to each other. Generating the image 61 enables the sound data to be represented as a two-dimensional image adapted to human hearing characteristics.

[0051] The image 61 is also illustrated in a rectangular shape like the image 60 (see Fig. 2A) but is not limited thereto. Modifiable matters in the image 60 also apply to the image 61.

[0052] The machine learning model generation unit 25 generates an optimum machine learning model for determining normality and abnormality in sound data using the image 61. Since the details are the same as those of the first embodiment, the description thereof is omitted.

[0053] Next, an operation example of the image processing apparatus 1 according to the second embodiment is described with reference to a flowchart of Fig. 7. However, the processes in steps S201 to S205 and S209 to S219 are the same as those in steps S101 to S105 and S107 to S117 in Fig. 4, and thus the description thereof is omitted.

[0054] In step S207, the scale setting unit 26 sets the twelve-tone scale and octaves that are overtones thereof using the "frequency-energy characteristics" acquired in step S203. The scale setting unit 26 classifies sounds of respective frequencies into the twelve-tone scale using the twelve-tone scale and octaves.

(Advantageous effects)

[0055] In the second embodiment, the fundamental frequency component and the harmonic component have the same scale. The second embodiment enables music-related events, such as a scale and an octave, to be displayed as an image.

[Third embodiment]

[0056] Next, a third embodiment of the present invention is described with reference to Fig. 8. The third embodiment differs from the first embodiment in that the controller 20 includes a critical band setting unit 27. A description of a configuration that overlaps the first embodiment is omitted with reference to reference numerals. The following description focuses on the difference.

[0057] The critical band setting unit 27 sets the twelve-tone scale of "C, C#, D, D#, E, F, G, G#, A, A#, B, B#, (C)", which corresponds to the "do, re, mi, fa, sol, la, ti (do)" used in music, and a critical band (band numbers are 1 to 24), which is human hearing characteristics, using the "frequency-energy characteristics" acquired by the frequency characteristics analysis unit 22. The critical

band setting unit 27 classifies sounds of respective frequencies into the twelve-tone scale using the twelve-tone scale and the critical band.

[0058] The critical band is defined as a maximum frequency band when the sound intensity of band noise with a constant band sound pressure level is constant regardless of a bandwidth. As another definition, the critical band is defined as a minimum frequency bandwidth of band noise, which becomes a sound pressure level at which a pure sound equal to a center frequency of the band noise is just heard when the bandwidth is increased while the spectrum level of the band noise is kept constant.

[0059] Next, an example of an image generated by the sound data image generation unit 24 is described with reference to Fig. 9.

[0060] An image 62 in Fig. 9 is an example of an image generated by the sound data image generation unit 24. The sound data image generation unit 24 generates the image 62 by arranging pixels indicating frequencies so as to correspond to the twelve-tone scale using the twelve-tone scale and critical band set by the critical band setting unit 27 and the frequencies classified by the critical band setting unit 27.

[0061] In Fig. 9, the horizontal axis indicates the twelve-tone scale, and the vertical axis indicates a critical band (band numbers are 1 to 24). In Fig. 9, the "C, C#, D, D#, E, F, G, G#, A, A#, B, B#, C" in Fig. 6 is omitted. In Fig. 9, the horizontal axis indicates 32 Hz to 130 Hz, which is a combination of a first octave and a second octave. The number of the critical band indicates the minimum frequency in that critical band. When the vertical axis is indicated by frequency, the frequency increases in the order of 100 Hz, 200 Hz, 300 Hz, 400 Hz, and 510 Hz. The reason why the increase in frequency is irregular on the vertical axis is because it simulates human hearing characteristics. Dashed-dotted lines in Fig. 9 indicate sensitivity characteristics with respect to the loudness of sound. A band surrounded by a dashed-dotted line indicates having a similar sensitivity characteristic.

[0062] In the third embodiment, the fundamental frequency is not 400 Hz but any value. As illustrated in Fig. 9, a pixel 60a indicating a fundamental frequency and a pixel 60b indicating a harmonic corresponding to the fundamental frequency are arranged adjacently. By generating the image 62 in which the fundamental frequency component and harmonic component having a relationship with respect to the critical band of human hearing are arranged adjacently, it is possible to represent sound data as a two-dimensional image matching human hearing characteristics.

[0063] The image 62 is also illustrated in a rectangular shape like the image 60 (see Fig. 2A) but is not limited thereto. Modifiable matters in the image 60 also apply to the image 62.

[0064] The machine learning model generation unit 25 generates an optimum machine learning model for determining normality and abnormality in sound data using the image 62. The details are the same as those of the first embodiment, and thus the description thereof is omitted.

[0065] Next, an operation example of the image processing apparatus 1 according to the third embodiment is described with reference to a flowchart of Fig. 10. However, the processes in steps S301 to S305 and S309 to S319 are the same as those in steps S101 to S105 and S107 to S117 in Fig. 4, and thus the description thereof is omitted.

[0066] In step S307, the critical band setting unit 27 sets the twelve-tone scale and the critical band using the "frequency-energy characteristics" acquired in step S303. The critical band setting unit 27 classifies sounds of respective frequencies into the twelve-tone scale using the twelve-tone scale and the critical band.

(Advantageous effects)

[0067] In the third embodiment, the fundamental frequency component and the harmonic component have a relationship with respect to the critical band of human hearing. The third embodiment enables such a relationship to be displayed as an image.

[Fourth embodiment]

[0068] Next, a fourth embodiment of the present invention is described with reference to Figs. 11 to 12. The fourth embodiment differs from the first embodiment in that the controller 20 includes a time-specific image generation unit 28 and a three-dimensional image generation unit 29 instead of the sound data image generation unit 24 (see Fig. 11). A description of a configuration that overlaps the first embodiment is omitted with reference to reference numerals. The following description focuses on the difference.

[0069] The image 60 in Fig. 12 is the same as that in Fig. 2A. The time-specific image generation unit 28 generates the image 60 using the fundamental frequency and harmonics calculated by the fundamental frequency calculation unit 23. Further, the time-specific image generation unit 28 performs the same processing every predetermined time to generate image 63 to 65. The three-dimensional image generation unit 29 generates a three-dimensional image 70 using the images 60, and 63 to 65 at every predetermined time generated by the time-specific image generation unit 28. Specifically, the three-dimensional image generation unit 29 adds a time axis to the images 60, and 63 to 65 (two-dimensional images) to generate the three-dimensional image 70. Modifiable matters in the image 60 also apply to the three-dimensional image 70. Due to calculation characteristics of the FFT, the data at every predetermined time are generally calculated by overlapping the data before and after the time axis, and a value at the predetermined time is calculated using data with a time length such that it has a predetermined frequency resolution.

[0070] The machine learning model generation unit 25 generates an optimum machine learning model for determining normality and abnormality in sound data using the three-dimensional image 70. The details are the same as those of the first embodiment, and thus the description thereof is omitted.

[0071] Next, an operation example of the image processing apparatus 1 according to the fourth embodiment is described with reference to a flowchart of Fig. 13. However, the processes in steps S401 to S405 and S413 to S419 are the same as those in steps S101 to S105 and S111 to S117 in Fig. 4, and thus the description thereof is omitted.

[0072] In step S407, the time-specific image generation unit 28 generates the images 60, and 63 to 65 at predetermined intervals using the fundamental frequency and harmonics acquired in step S405. The process proceeds to step S409, and the three-dimensional image generation unit 29 generates the three-dimensional image 70 using the images 60, and 63 to 65 generated at every predetermined time in step S407. The process proceeds to step S411, and the three-dimensional image 70 generated in step S409 is displayed on the display 50. Each pixel forming the three-dimensional image 70 is set as a brightness or color corresponding to the amplitude of the sound data. Thus, the operator who sees the three-dimensional image 70 can grasp the intensity, normality, abnormality, and the like of the sound data at a glance.

(Advantageous effects)

[0073] The fourth embodiment enables sound data to be displayed as a three-dimensional image.

[Fifth embodiment]

[0074] Next, a fifth embodiment of the present invention is described with reference to Figs. 14 to 15. The fifth embodiment differs from the fourth embodiment in that the controller 20 further includes a frequency setting unit 30 and an image cutout unit 31 (see Fig. 14). A description of the configuration that overlaps the fourth embodiment is omitted with reference to reference numerals. The following description focuses on the difference.

[0075] The frequency setting unit 30 sets a frequency to be extracted from the three-dimensional image 70. The frequency set by the frequency setting unit 30 is any frequency, and the fundamental frequency may be set.

[0076] The image cutout unit 31 cuts out pixels related to the frequency set by the frequency setting unit 30. Specifically, as illustrated in Fig. 15, the image cutout unit 31 cuts out data 80 of a plurality of pixels corresponding to the frequency set by the frequency setting unit 30. The image cutout unit 31 generates a spectrogram using the cutout data 80. A spectrogram is an arrangement in which a spectrum of frequency characteristics is arranged on a time axis, and intensity is represented by colors, like a heat map.

[0077] The machine learning model generation unit 25 generates an optimum machine learning model for determining normality and abnormality in sound data using the spectrogram. The details are the same as those of the first embodiment, and thus the description thereof is omitted.

[0078] Next, an operation example of the image processing apparatus 1 according to the fifth embodiment is described with reference to a flowchart of Fig. 16. However, the processes in steps S501 to S509 and S515 to S521 are the same as those in steps S401 to S409 and S413 to S419 in Fig. 13, the description thereof is omitted.

[0079] In step S511, the frequency setting unit 30 sets a frequency to be extracted from the three-dimensional image 70. The process proceeds to step S513, and the image cutout unit 31 cuts out pixels relating to the frequency set in step S511. The image cutout unit 31 generates a spectrogram using the cutout pixels.

(Advantageous effects)

[0080] The fifth embodiment enables analysis to be performed using a spectrogram.

[Sixth embodiment]

[0081] Next, a sixth embodiment of the present invention is described with reference to Fig. 17. The sixth embodiment differs from the first embodiment in that the controller 20 includes a new sound data acquisition unit 32, an image processing unit 33, and a determination unit 34. A description of a configuration that overlaps the first embodiment is omitted with reference to reference numerals. The following description focuses on the difference.

[0082] The new sound data acquisition unit 32 acquires new sound data via a microphone 11 different from the microphone 10. The microphones 10 and 11 are attached to machines of the same type. The sound data image generation unit 24 generates an image of the new sound data.

[0083] The image processing unit 33 uses the image of the new sound data as input data of the machine learning model generated by the machine learning model generation unit 25. The image processing unit 33 outputs an index, such as the degree of agreement of images, using a predetermined image processing method.

[0084] The determination unit 34 compares the output value output by the image processing unit 33 with a preset threshold value. When the output value is greater than the threshold value, the determination unit 34 determines that the sound data is normal. In contrast, when the output value is equal to or less than the threshold value, the determination unit 34 determines that the sound data is abnormal. Note that the determination method is not limited to a method using a threshold value, and other methods may be used.

[0085] The determination result by the determination unit 34 is displayed on the display 50. Although not illustrated, the determination result by the determination unit 34 may be notified by voice through a speaker. When the determination result by the determination unit 34 is abnormal, a red rotating light may be turned on.

[0086] Next, an operation example of the image processing apparatus 1 according to the sixth embodiment is described with reference to flowcharts of Figs. 18A and 18B. However, the processes in step S601 to S613 are the same as those in steps S101 to S113 in Fig. 4, and thus the description thereof is omitted.

[0087] In step S615, the new sound data acquisition unit 32 acquires new sound data via the microphone 11. The acquired new sound data is converted into an electric signal and treated as time series data. The process proceeds to step S617, and the frequency characteristics analysis unit 22 analyzes the electric signal acquired in step S615 using the FFT. The FFT provides the "frequency-energy characteristics".

[0088] The process proceeds to step S619, and the fundamental frequency calculation unit 23 calculates a fundamental frequency of the new sound data using the "frequency-energy characteristics" acquired in step S617. The process proceeds to step S621, and it is determined whether the fundamental frequency calculated in step S619 agrees with the fundamental frequency of the machine learning model generated in step S613. The reason for this determination is that when the fundamental frequencies are different, the machine learning model cannot perform the normal and abnormal determination processing. Note that "the fundamental frequencies agree with each other" means substantial agreement.

[0089] When the fundamental frequencies do not agree with each other (NO in step S621), "Determination processing not possible due to disagreement of fundamental frequencies" is displayed on the display 50, and the process proceeds to step S631. In contrast, when the fundamental frequencies agree with each other (YES in step S621), the process proceeds to step S623, and the sound data image generation unit 24 generates an image of the new sound data using the fundamental frequency calculated in step S619.

[0090] The process proceeds to step S625, and the image processing unit 33 uses the image of the new sound data generated in step S623 as input data for the machine learning model. The image processing unit 33 outputs an index, such as the degree of agreement of images, using a predetermined image processing method. The determination unit 34 compares the output value output by the image processing unit 33 with a preset threshold value to determine whether the sound data is normal or abnormal.

[0091] The process proceeds to step S627, and the determination result of step S625 is displayed on the display 50. The process proceeds to step S629, and a file name of the new sound data, a name of the machine learning model, a processing execution time, a value of the fundamental frequency, a determination result, and the like are stored in the storage device 40. A series of processing is repeatedly executed until completion (step S631). Note that when the processing is completed, a notice "End of normal/abnormal determination processing" may be displayed on the display 50.

(Advantageous effects)

[0092] The sixth embodiment makes it possible to determine whether other sound data are normal or abnormal using a machine learning model acquired by imaging sound data.

[0093] Each of the functions described in the above embodiments may be implemented by one or more processing circuits. The processing circuit includes a programmed processing device, such as a processing device including an electric circuit. The processing circuit also includes devices, such as an application specific integrated circuit (ASIC) arranged to perform the described functions and circuit components.

[0094] While embodiments of the present invention have been described as above, the statements and drawings that form part of this disclosure should not be understood as limiting the invention. Various alternative embodiments, examples, and operating techniques will become apparent to those skilled in the art from this disclosure.

[0095] In the above-described embodiments, a machine learning model is used as a method for determining normality and abnormality in other sound data, but the method is not limited thereto. For example, an abnormality in the image (sound data) may be determined by comparing the fundamental frequency component and harmonic component with other frequency components. This makes it possible to determine whether the sound data is normal or abnormal in a case where there is no overtone relation, such as a critical band.

[0096] Further, the determination unit 34 may determine an abnormality in a predetermined sound included in sound data using the image 60 (sound image).

[0097] The image 60 (sound image) may be made from a two-dimensional matrix including a fundamental frequency component and harmonic component converted into image data, and other frequency components converted into image data, wherein a predetermined area is set for each frequency component. Note that the other frequency components mean frequency components other than the fundamental frequency component and the harmonic component.

REFERENCE SIGNS LIST

[0098]

1 Image processing apparatus
20 Controller
21 Sound data acquisition unit

22 Frequency characteristics analysis unit
23 Fundamental frequency calculation unit
24 Sound data image generation unit
25 Machine learning model generation unit
26 Scale setting unit
27 Critical band setting unit
28 Time-specific image generation unit
30 Frequency setting unit
31 Image cutout unit
32 New sound data acquisition unit
33 Image processing unit
34 Determination unit
40 Storage device
50 Display

Claims

1. An image processing apparatus for acquiring sound data and analyzing the sound data that have been acquired, the image processing apparatus comprising:

a controller configured to: calculate a fundamental frequency component included in the sound data and a harmonic component corresponding to the fundamental frequency component;
convert the fundamental frequency component and the harmonic component that have been calculated into image data; and
generate a sound image in which the fundamental frequency component and the harmonic component that have been converted into the image data are arranged adjacent each other.

2. The image processing apparatus according to claim 1, wherein the fundamental frequency component and the harmonic component have an overtone relationship, a same pitch name, or a relationship with respect to a critical band of human hearing.

3. The image processing apparatus according to claim 1 or 2, wherein the controller arranges the fundamental frequency component and the harmonic component adjacent to each other on a vertical axis or a horizontal axis of the sound image.

4. The image processing apparatus according to any one of claims 1 to 3, wherein the controller generates the sound image by converting the fundamental frequency component and the harmonic component into image data in which a brightness or a color corresponding to an amplitude of the sound data is set.

5. The image processing apparatus according to claim 1, wherein the controller arranges a plurality of frequency components of the sound data in frequency order on a vertical axis or a horizontal axis of the sound image.

6. The image processing apparatus according to claim 1, wherein the sound image includes a two-dimensional matrix including the fundamental frequency component and the harmonic component that have been converted into the image data, and other frequency components that have been converted into image data, the two-dimensional matrix having a predetermined area set for each frequency component.

7. The image processing apparatus according to any one of claims 1 to 6, wherein the controller determines an abnormality in a predetermined sound included in the sound data using the sound image.

8. The image processing apparatus according to any one of claims 1 to 6, wherein the controller generates a machine learning model by machine learning the sound image and determines an abnormality in the sound image using the machine learning model.

9. The image processing apparatus according to claim 1, wherein the sound image includes a two-dimensional matrix including the fundamental frequency component and the harmonic component that have been converted into the image data, and other frequency components that have been converted into image data, the two-dimensional matrix having a predetermined area set for each frequency component, and
the controller determines an abnormality in the sound image by comparing the fundamental frequency component and the harmonic component with the other frequency components.

10. A method of processing an image, comprising:

acquiring sound data;
calculating a fundamental frequency component included in the sound data and a harmonic component corresponding to the fundamental frequency component;
converting the fundamental frequency component and the harmonic component that have been calculated into image data; and
generating a sound image where the fundamental frequency component and the harmonic component that have been converted into the image data are arranged adjacent each other.

FIG. 1

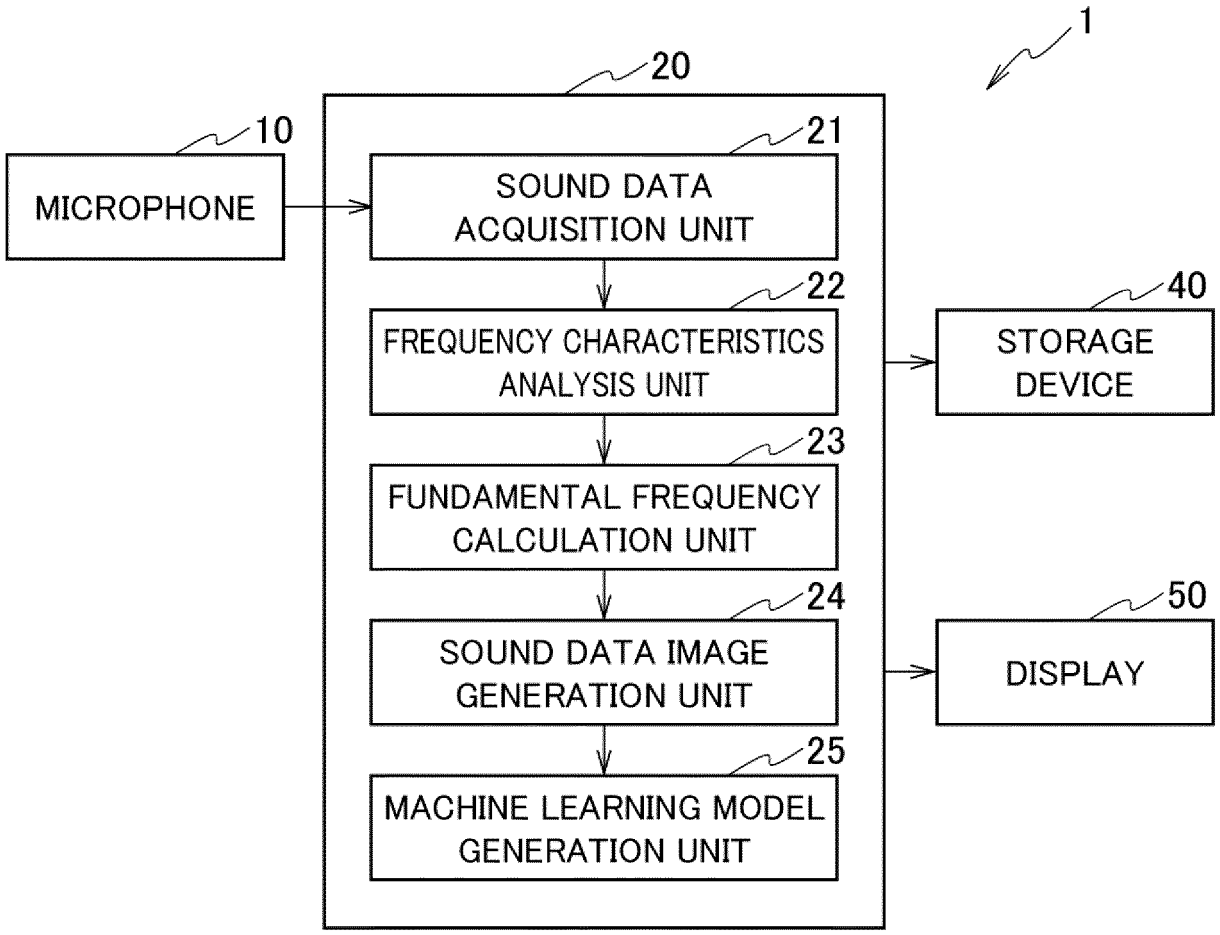

1
20
10
MICROPHONE
21
SOUND DATA ACQUISITION UNIT
22
FREQUENCY CHARACTERISTICS ANALYSIS UNIT
23
FUNDAMENTAL FREQUENCY CALCULATION UNIT
24
SOUND DATA IMAGE GENERATION UNIT
25
MACHINE LEARNING MODEL GENERATION UNIT
40
STORAGE DEVICE
50
DISPLAY

FIG. 2A

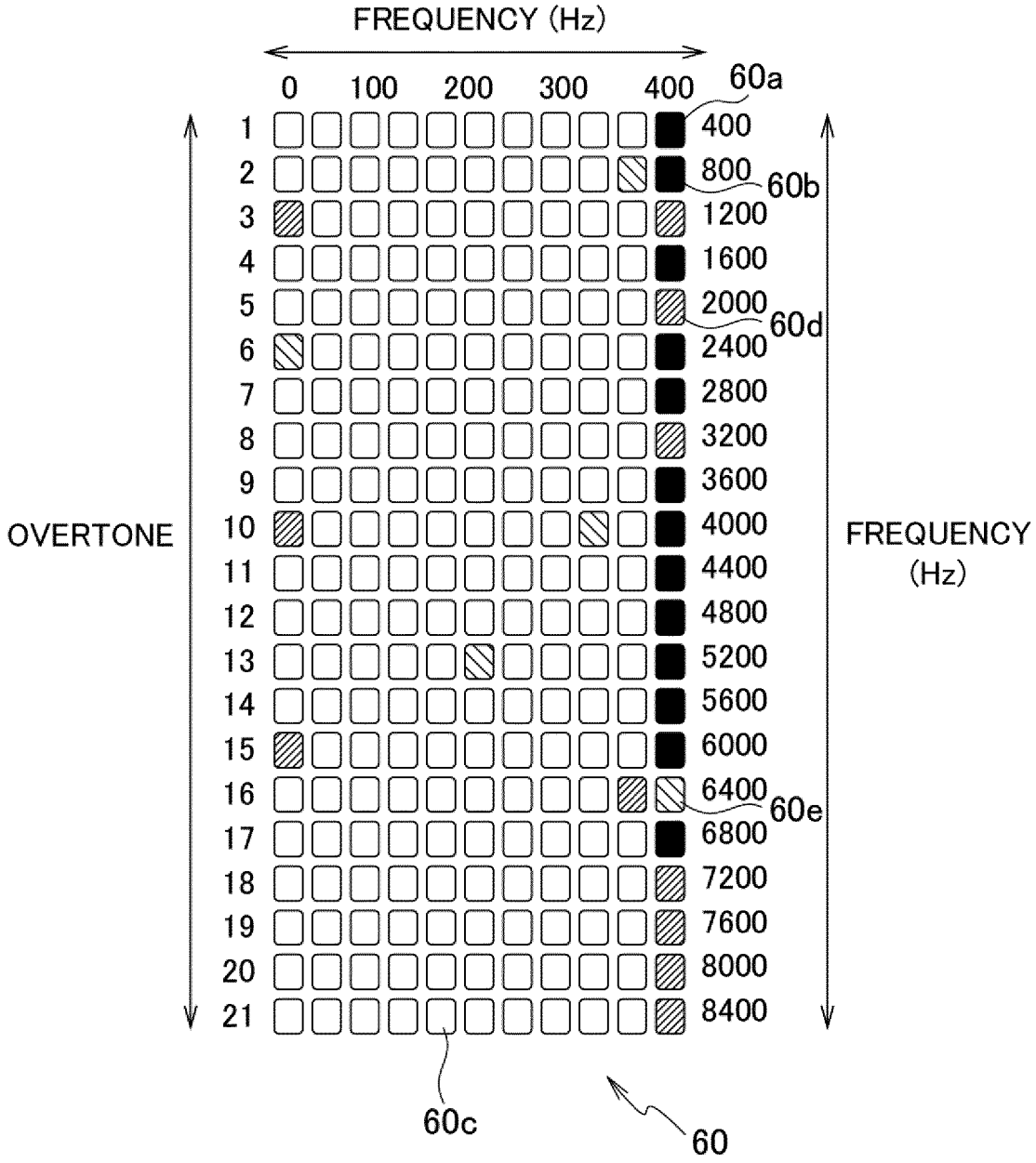

FREQUENCY (Hz)
0
100
200
300
400
60a
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
400
800
1200
1600
2000
2400
2800
3200
3600
4000
4400
4800
5200
5600
6000
6400
6800
7200
7600
8000
8400
60b
60d
60e
OVERTONE
FREQUENCY
(Hz)
60c
60

FIG. 2B

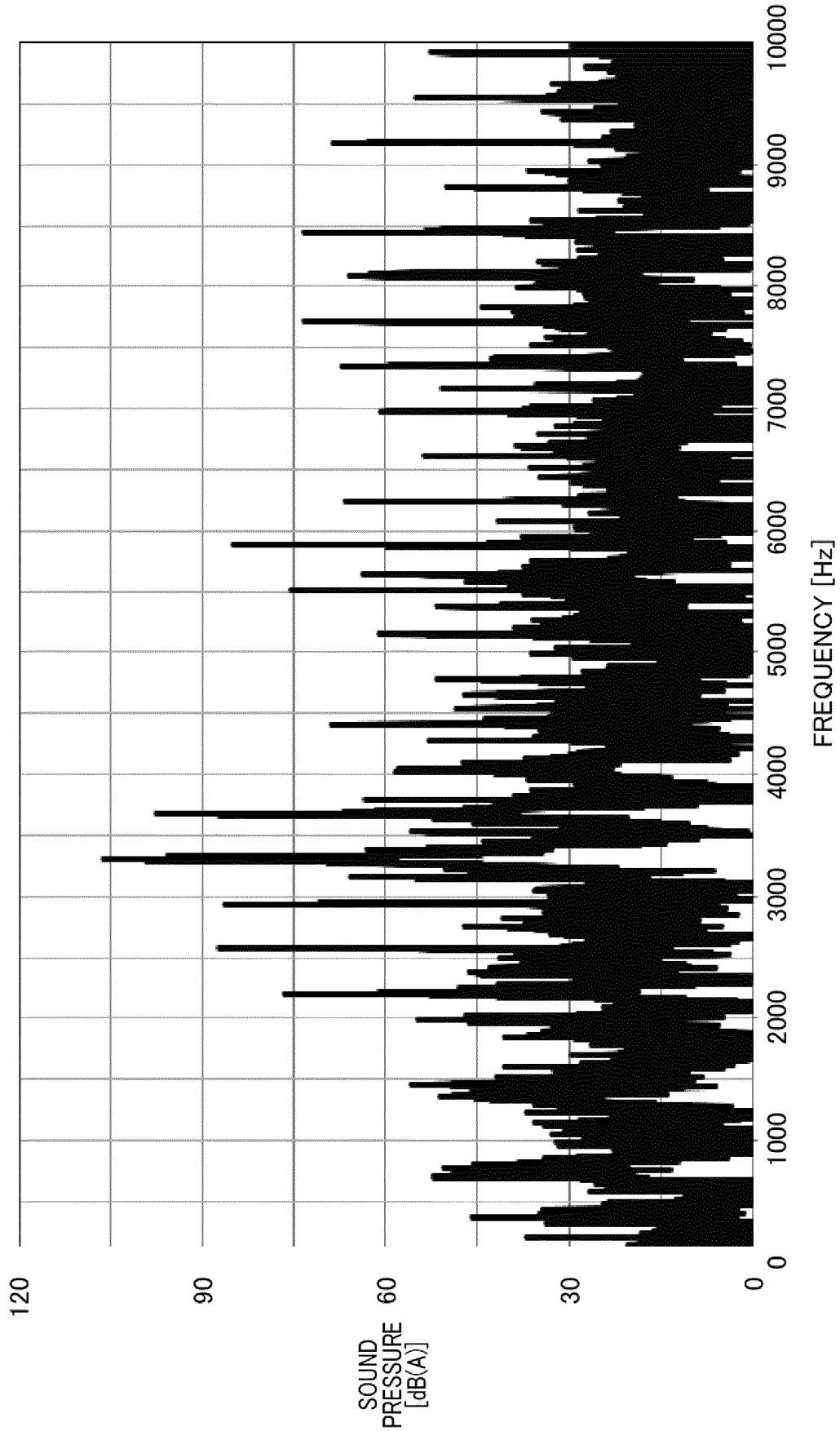
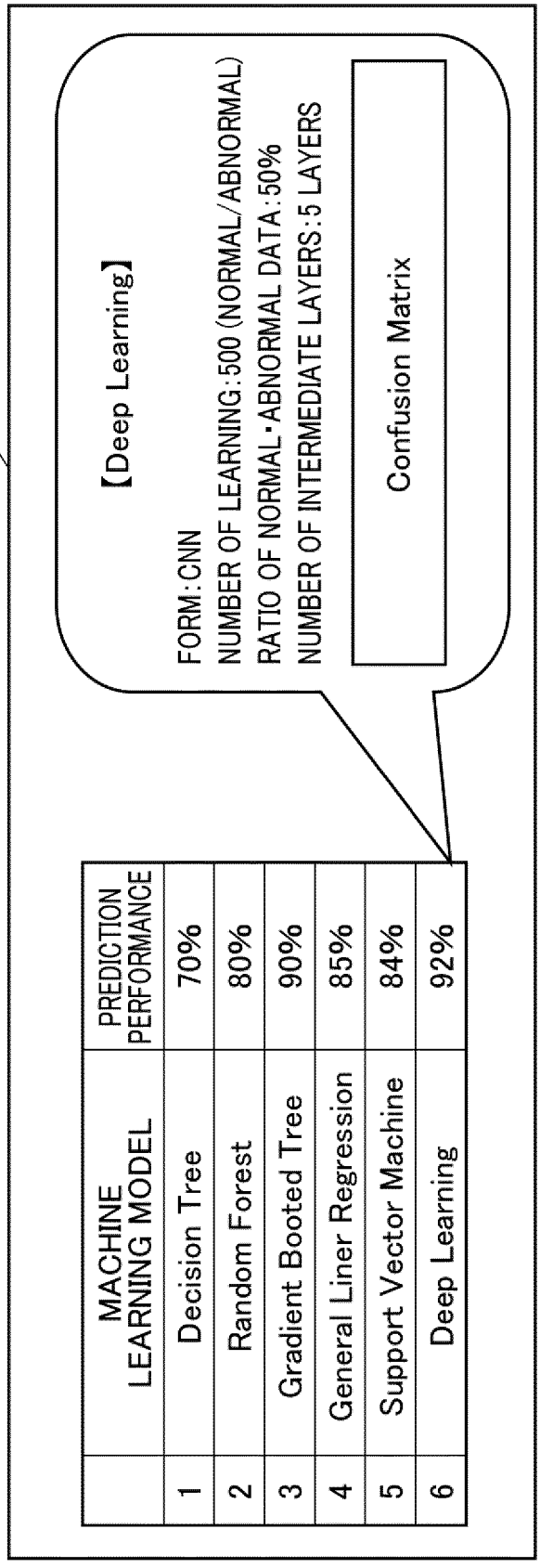

SOUND PRESSURE [dB(A)]
120
90
60
30
0
FREQUENCY [Hz]
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000

FIG. 3



	MACHINE LEARNING MODEL	PREDICTION PERFORMANCE
1	Decision Tree	70%
2	Random Forest	80%
3	Gradient Booted Tree	90%
4	General Liner Regression	85%
5	Support Vector Machine	84%
6	Deep Learning	92%

FIG. 4

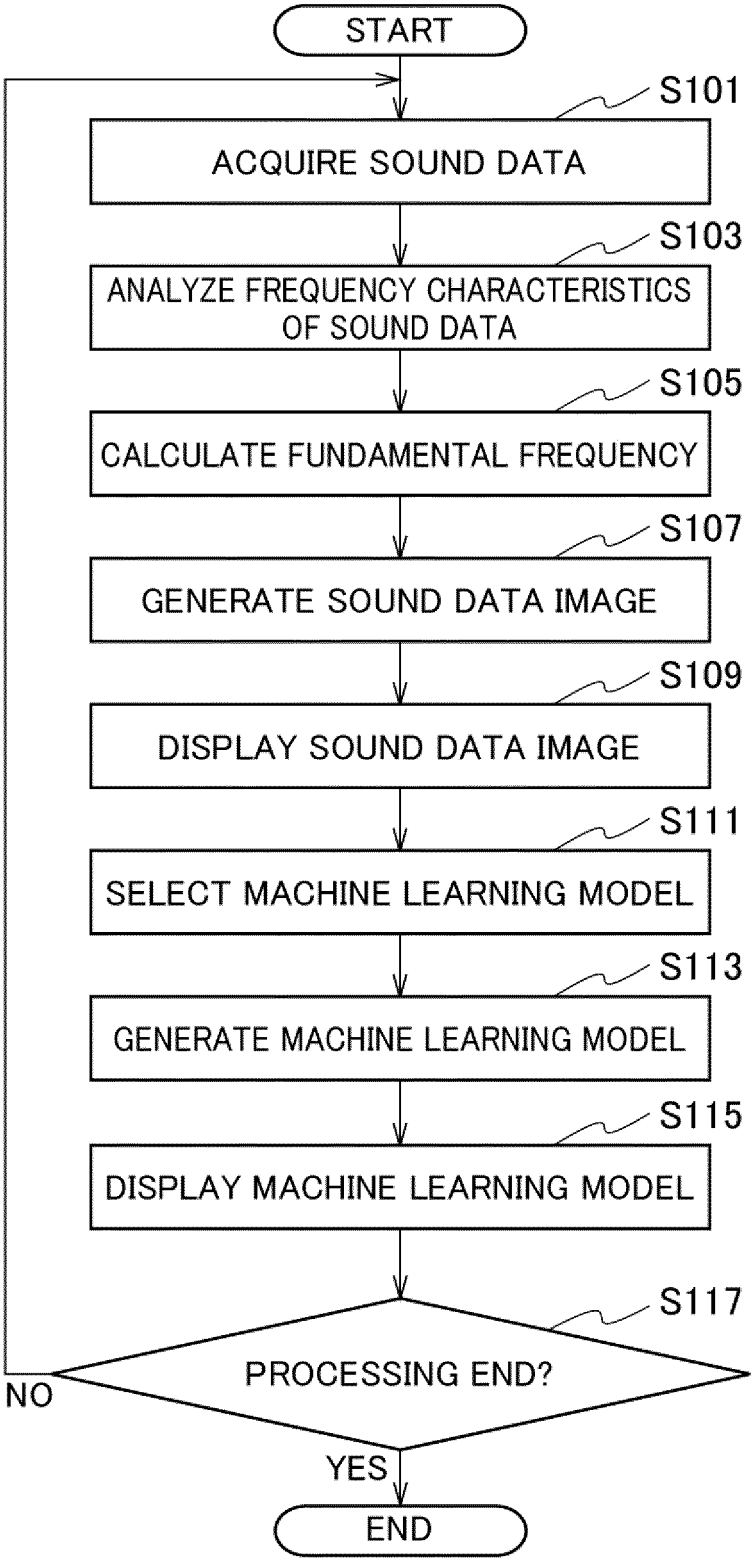

START
S101
ACQUIRE SOUND DATA
S103
ANALYZE FREQUENCY CHARACTERISTICS OF SOUND DATA
S105
CALCULATE FUNDAMENTAL FREQUENCY
S107
GENERATE SOUND DATA IMAGE
S109
DISPLAY SOUND DATA IMAGE
S111
SELECT MACHINE LEARNING MODEL
S113
GENERATE MACHINE LEARNING MODEL
S115
DISPLAY MACHINE LEARNING MODEL
S117
PROCESSING END?
NO
YES
END

FIG. 5

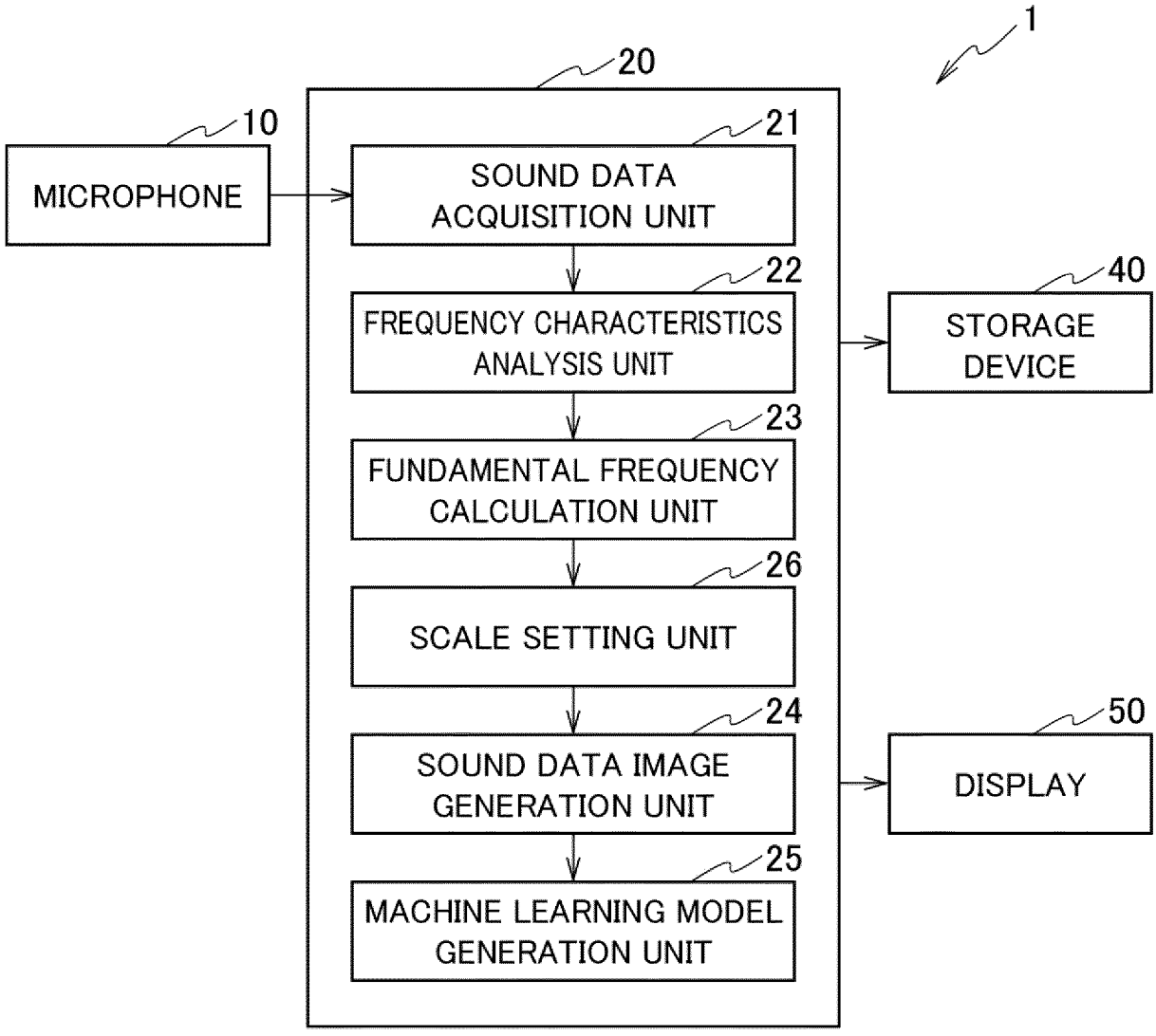

1
20
10
MICROPHONE
21
SOUND DATA ACQUISITION UNIT
22
FREQUENCY CHARACTERISTICS ANALYSIS UNIT
23
FUNDAMENTAL FREQUENCY CALCULATION UNIT
26
SCALE SETTING UNIT
24
SOUND DATA IMAGE GENERATION UNIT
25
MACHINE LEARNING MODEL GENERATION UNIT
40
STORAGE DEVICE
50
DISPLAY

FIG. 6

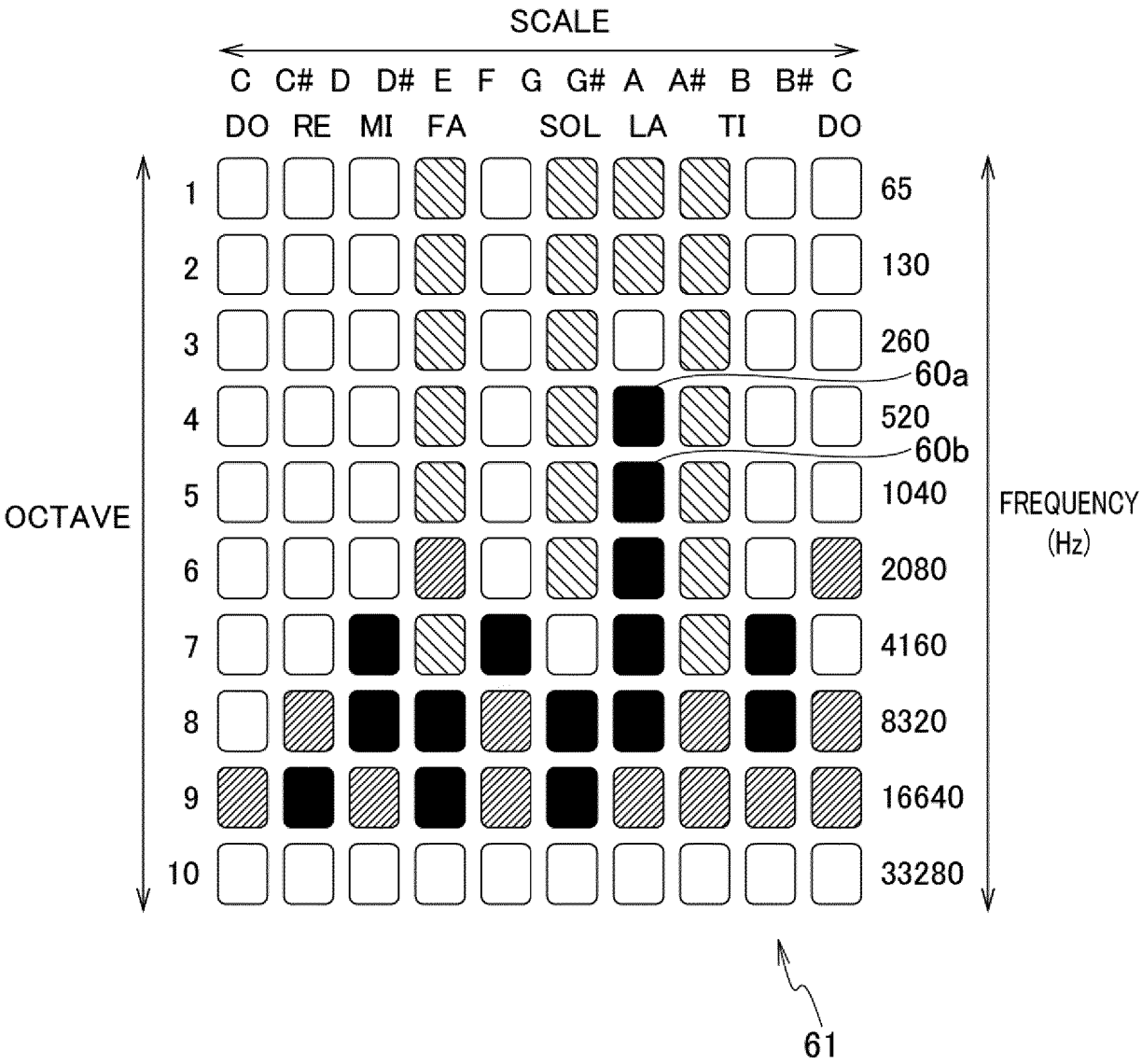

SCALE
C C# D D# E F G G# A A# B B# C
DO RE MI FA SOL LA TI DO
OCTAVE
1
2
3
4
5
6
7
8
9
10
65
130
260
520
1040
2080
4160
8320
16640
33280
FREQUENCY (Hz)
60a
60b
61

FIG. 7

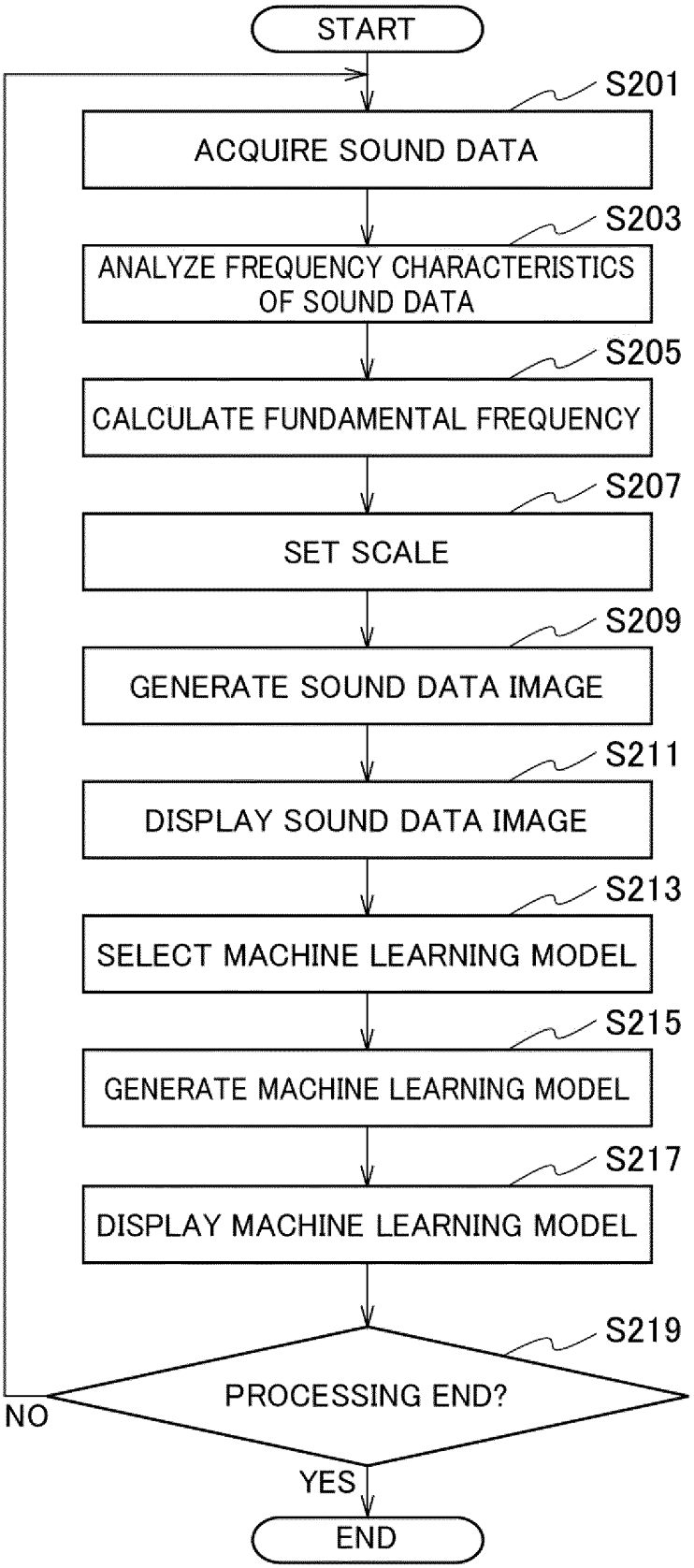

START
S201
ACQUIRE SOUND DATA
S203
ANALYZE FREQUENCY CHARACTERISTICS OF SOUND DATA
S205
CALCULATE FUNDAMENTAL FREQUENCY
S207
SET SCALE
S209
GENERATE SOUND DATA IMAGE
S211
DISPLAY SOUND DATA IMAGE
S213
SELECT MACHINE LEARNING MODEL
S215
GENERATE MACHINE LEARNING MODEL
S217
DISPLAY MACHINE LEARNING MODEL
S219
PROCESSING END?
NO
YES
END

FIG. 8

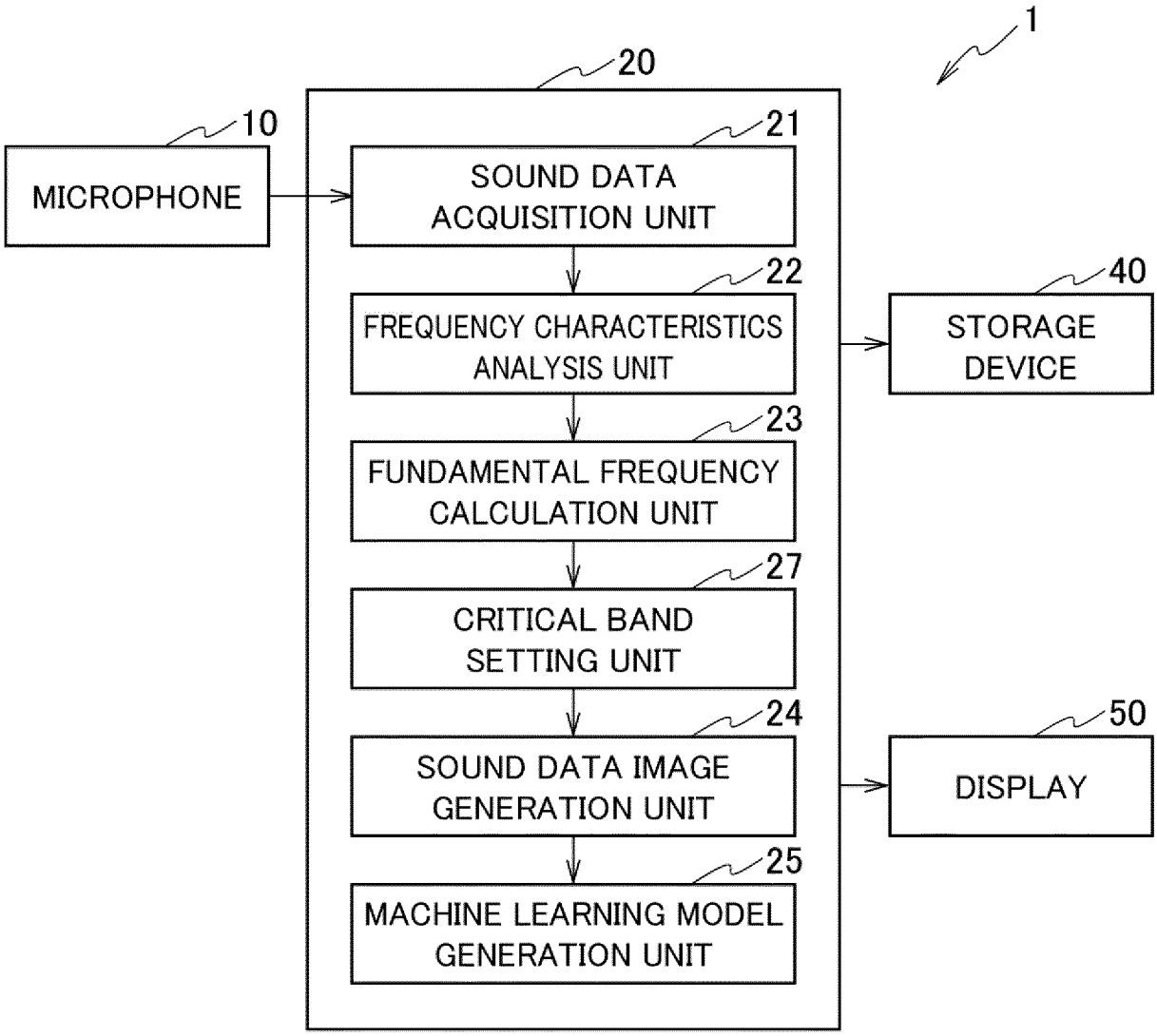

1
20
10
MICROPHONE
21
SOUND DATA ACQUISITION UNIT
22
FREQUENCY CHARACTERISTICS ANALYSIS UNIT
23
FUNDAMENTAL FREQUENCY CALCULATION UNIT
27
CRITICAL BAND SETTING UNIT
24
SOUND DATA IMAGE GENERATION UNIT
25
MACHINE LEARNING MODEL GENERATION UNIT
40
STORAGE DEVICE
50
DISPLAY

FIG. 9

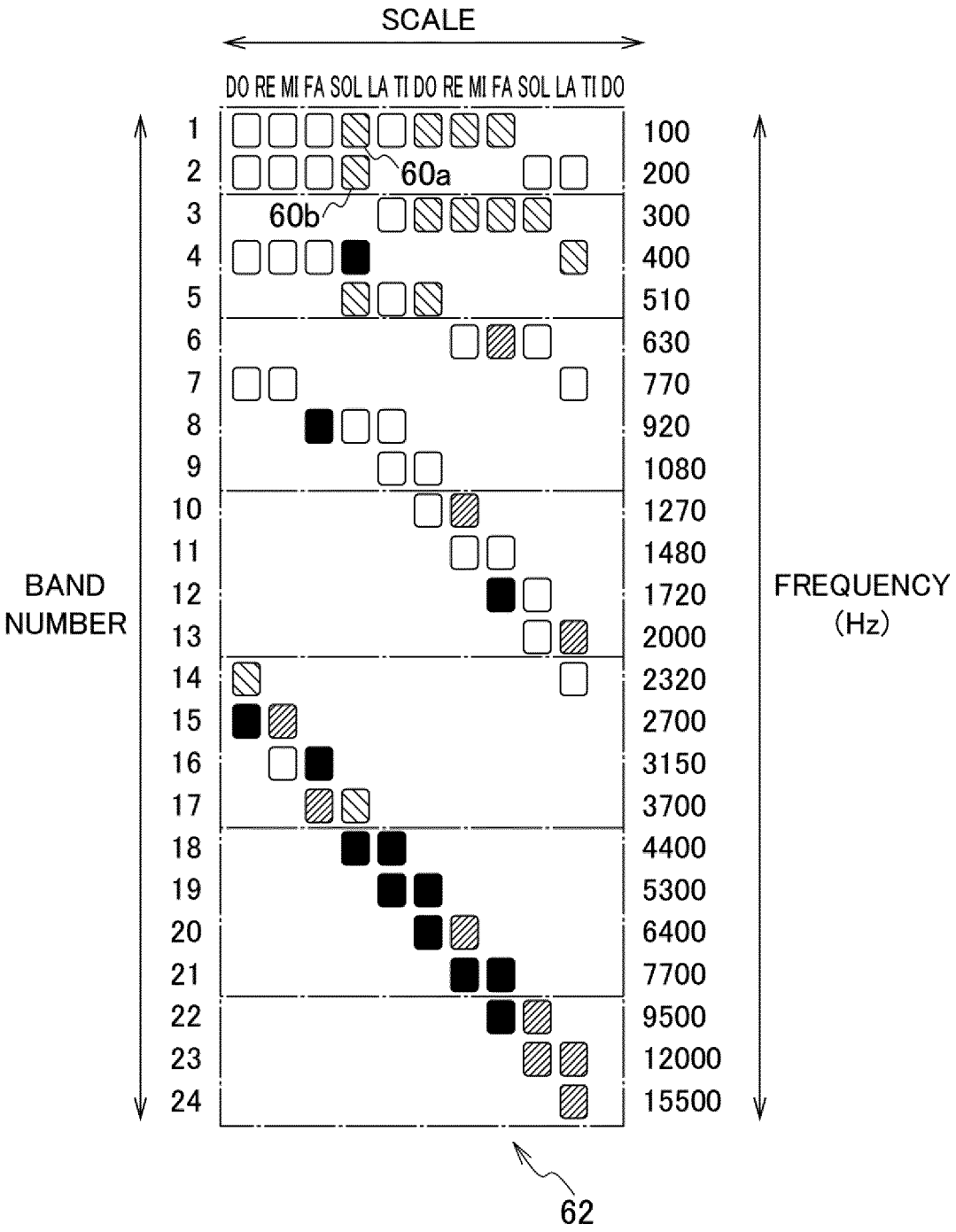

SCALE
DO RE MI FA SOL LA TI DO RE MI FA SOL LA TI DO
BAND NUMBER
FREQUENCY (Hz)
1 100
2 200
3 300
4 400
5 510
6 630
7 770
8 920
9 1080
10 1270
11 1480
12 1720
13 2000
14 2320
15 2700
16 3150
17 3700
18 4400
19 5300
20 6400
21 7700
22 9500
23 12000
24 15500
60a
60b
62

FIG. 10

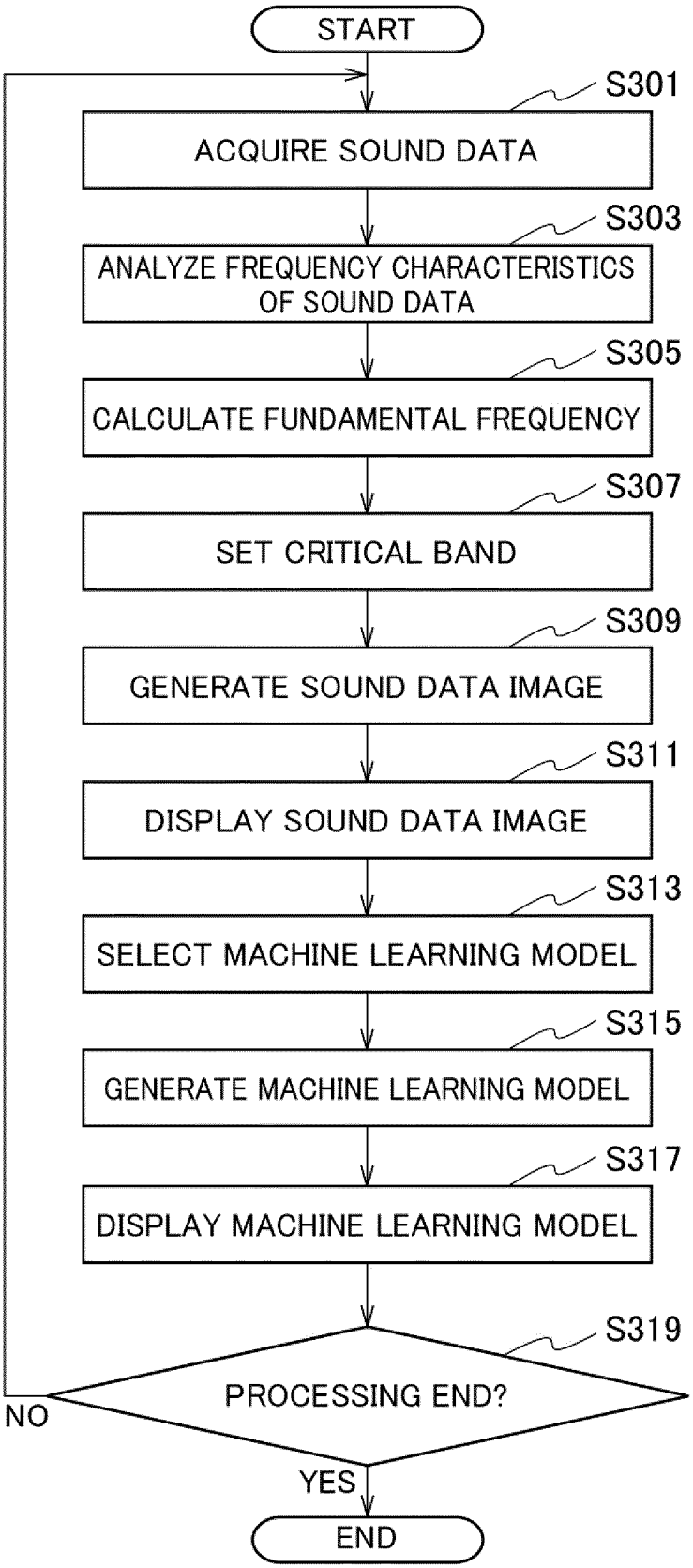

START
S301
ACQUIRE SOUND DATA
S303
ANALYZE FREQUENCY CHARACTERISTICS OF SOUND DATA
S305
CALCULATE FUNDAMENTAL FREQUENCY
S307
SET CRITICAL BAND
S309
GENERATE SOUND DATA IMAGE
S311
DISPLAY SOUND DATA IMAGE
S313
SELECT MACHINE LEARNING MODEL
S315
GENERATE MACHINE LEARNING MODEL
S317
DISPLAY MACHINE LEARNING MODEL
S319
PROCESSING END?
NO
YES
END

FIG. 11

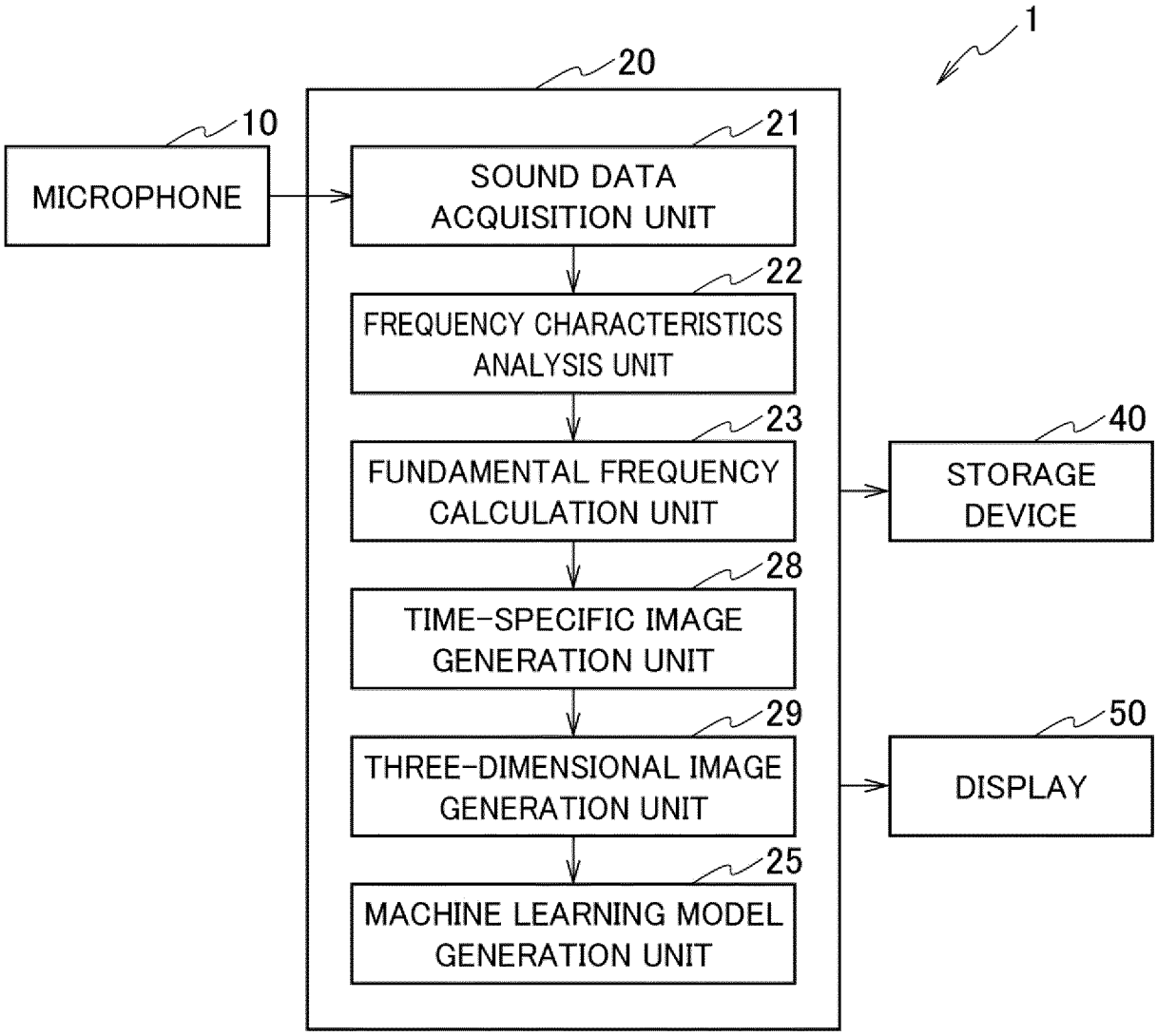

1
20
10
MICROPHONE
21
SOUND DATA ACQUISITION UNIT
22
FREQUENCY CHARACTERISTICS ANALYSIS UNIT
23
FUNDAMENTAL FREQUENCY CALCULATION UNIT
28
TIME-SPECIFIC IMAGE GENERATION UNIT
29
THREE-DIMENSIONAL IMAGE GENERATION UNIT
25
MACHINE LEARNING MODEL GENERATION UNIT
40
STORAGE DEVICE
50
DISPLAY

FIG. 12

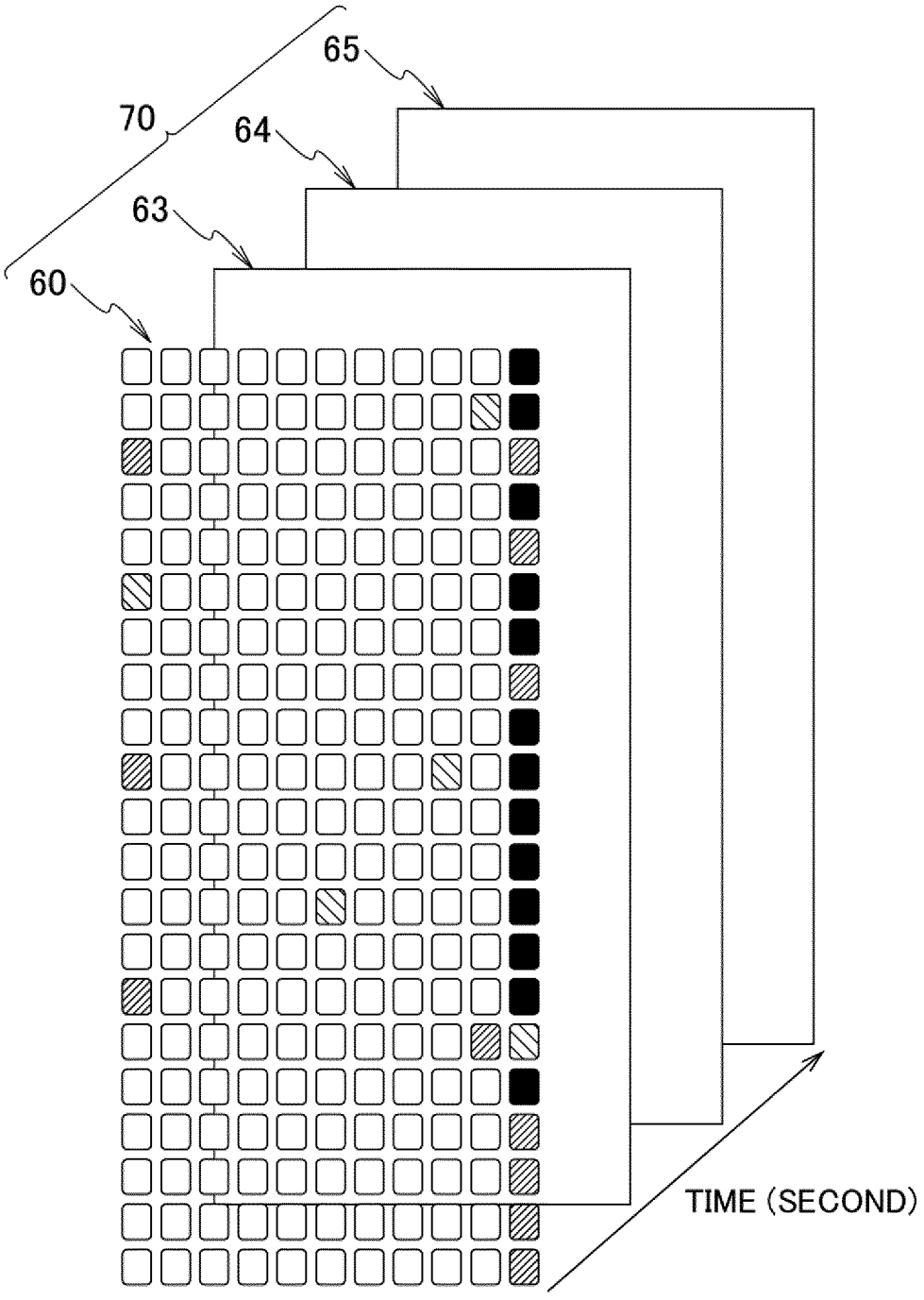


FIG. 13

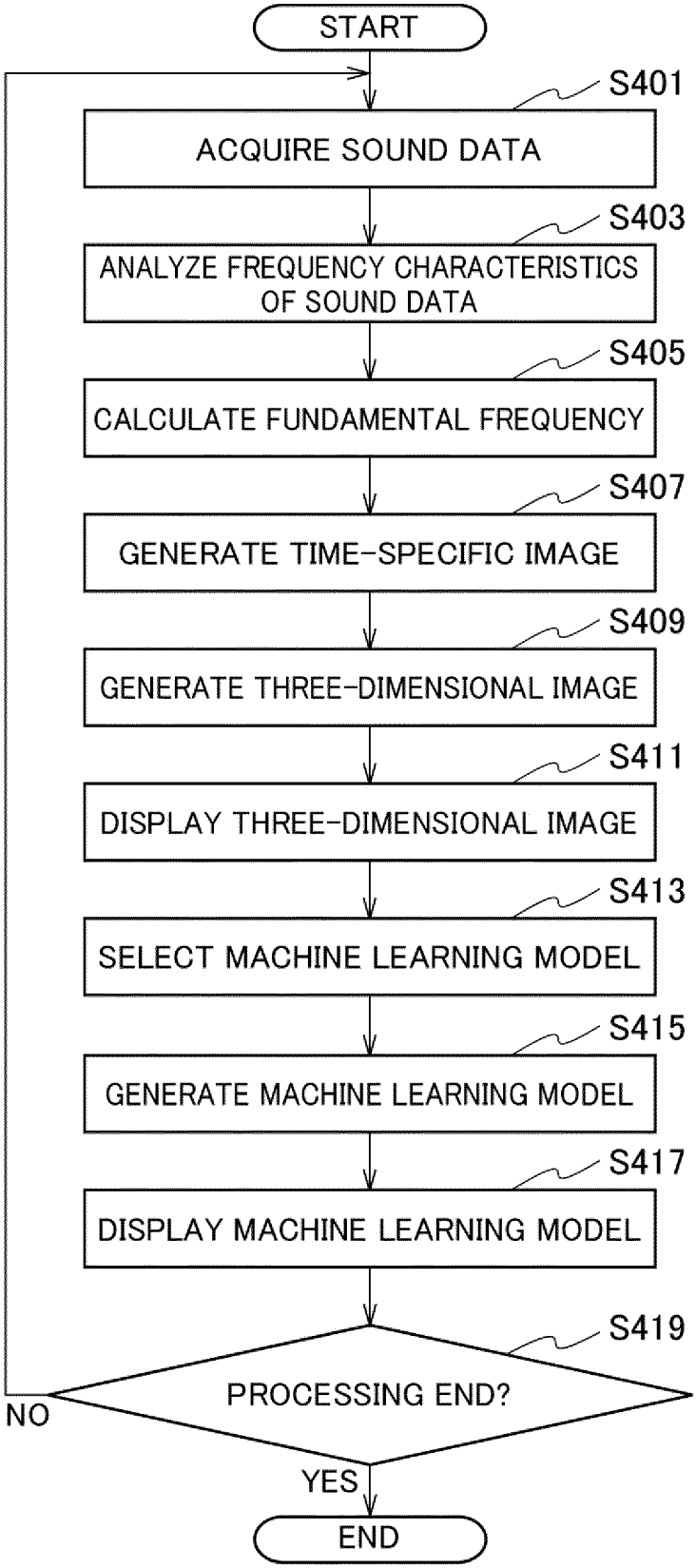

START
S401
ACQUIRE SOUND DATA
S403
ANALYZE FREQUENCY CHARACTERISTICS OF SOUND DATA
S405
CALCULATE FUNDAMENTAL FREQUENCY
S407
GENERATE TIME-SPECIFIC IMAGE
S409
GENERATE THREE-DIMENSIONAL IMAGE
S411
DISPLAY THREE-DIMENSIONAL IMAGE
S413
SELECT MACHINE LEARNING MODEL
S415
GENERATE MACHINE LEARNING MODEL
S417
DISPLAY MACHINE LEARNING MODEL
S419
PROCESSING END?
NO
YES
END

FIG. 14

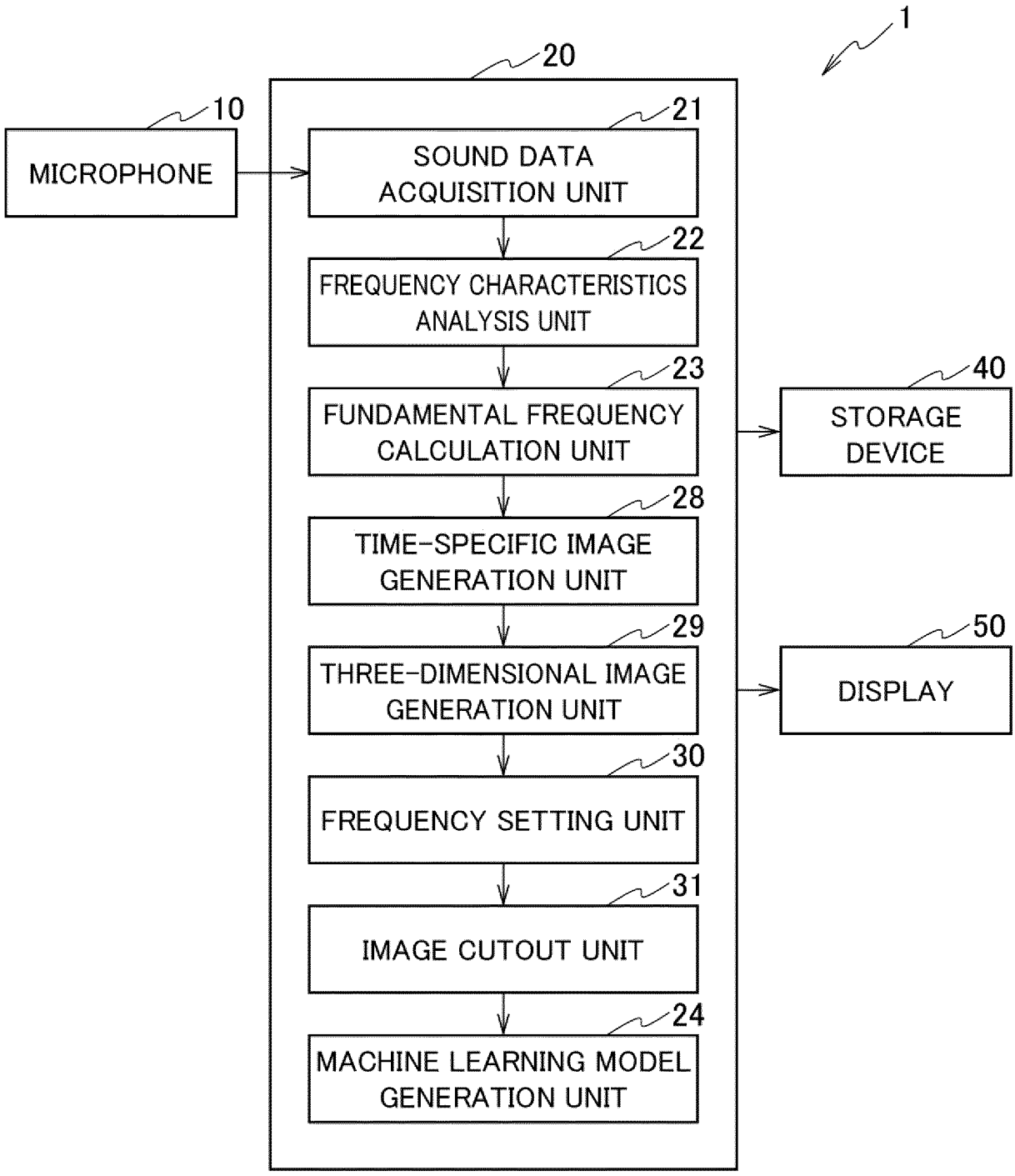

1
20
10
MICROPHONE
21
SOUND DATA ACQUISITION UNIT
22
FREQUENCY CHARACTERISTICS ANALYSIS UNIT
23
FUNDAMENTAL FREQUENCY CALCULATION UNIT
28
TIME-SPECIFIC IMAGE GENERATION UNIT
29
THREE-DIMENSIONAL IMAGE GENERATION UNIT
30
FREQUENCY SETTING UNIT
31
IMAGE CUTOUT UNIT
24
MACHINE LEARNING MODEL GENERATION UNIT
40
STORAGE DEVICE
50
DISPLAY

FIG. 15

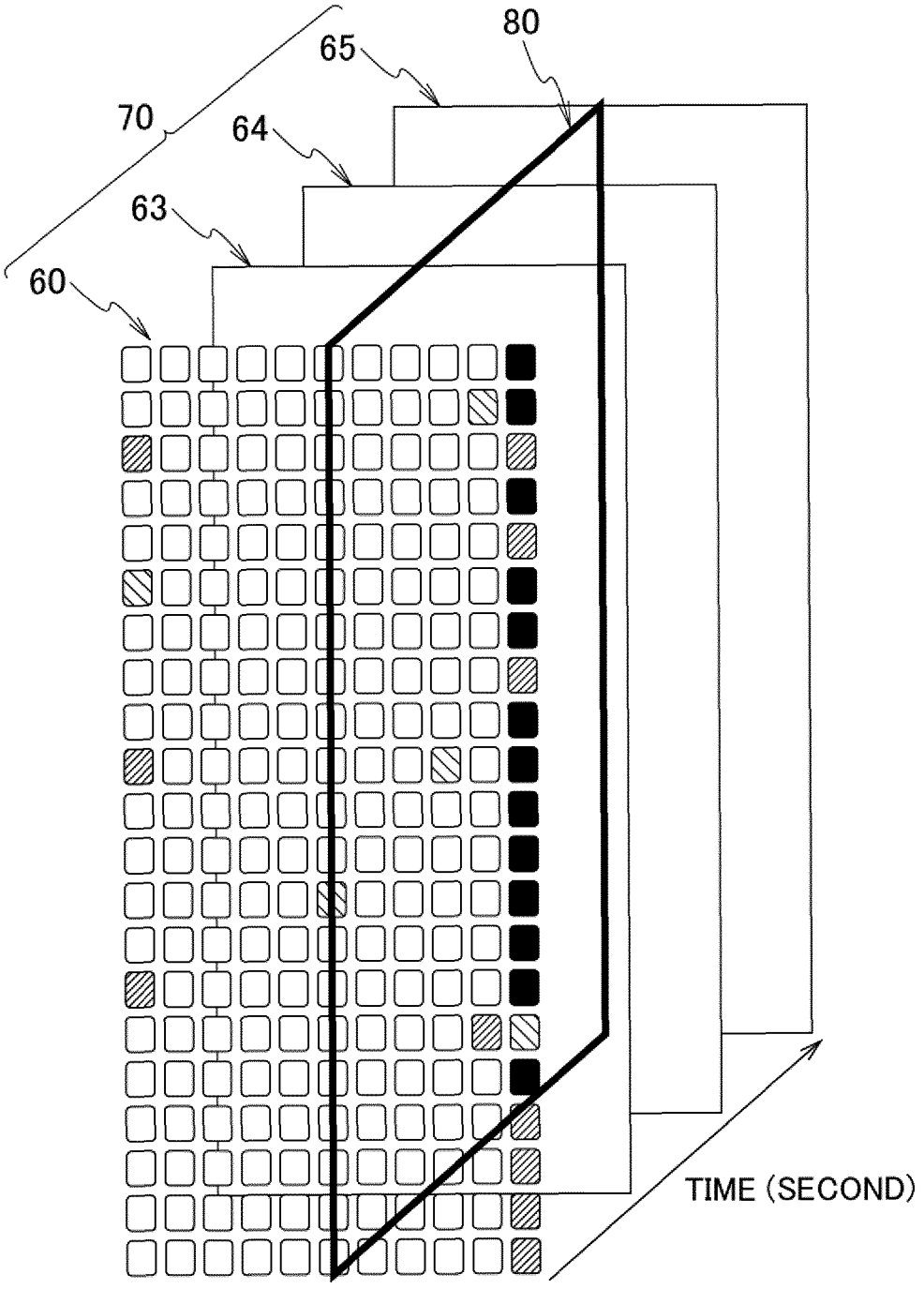


FIG. 16

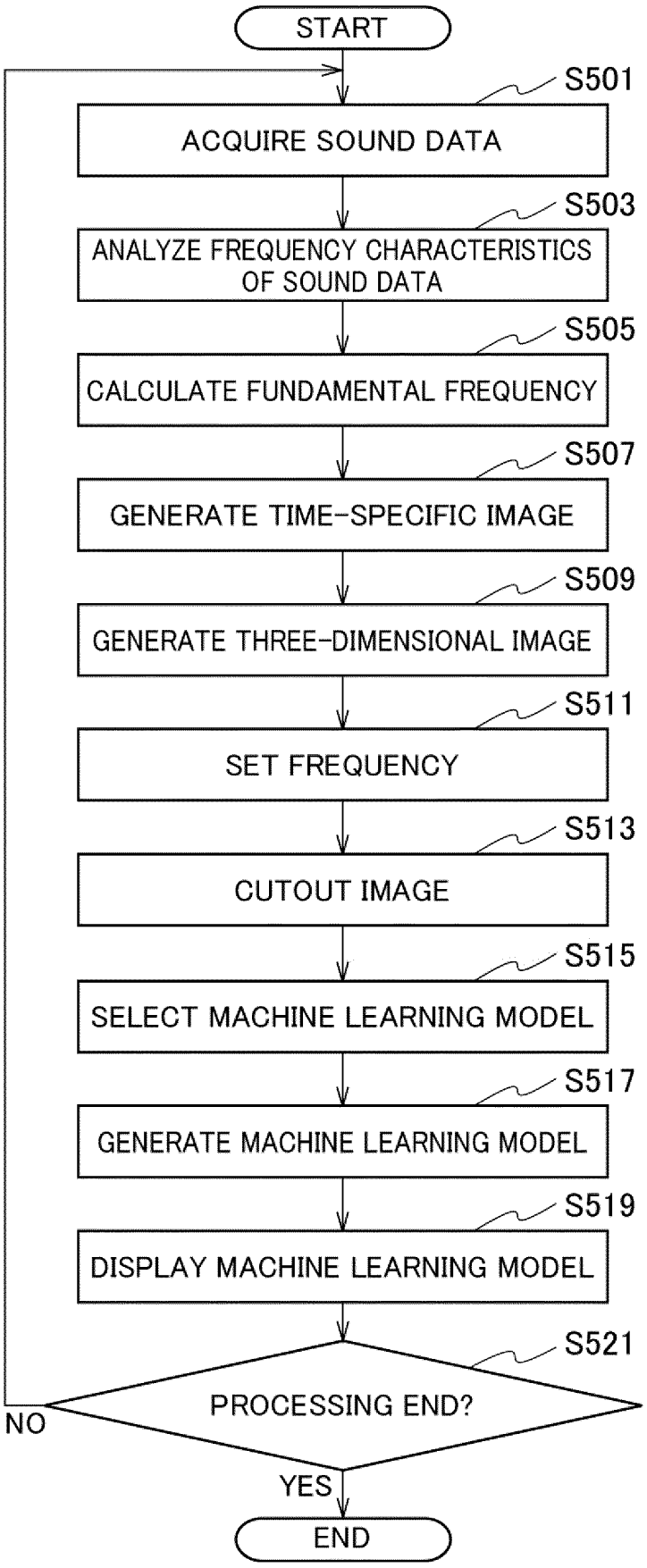

START
S501
ACQUIRE SOUND DATA
S503
ANALYZE FREQUENCY CHARACTERISTICS OF SOUND DATA
S505
CALCULATE FUNDAMENTAL FREQUENCY
S507
GENERATE TIME-SPECIFIC IMAGE
S509
GENERATE THREE-DIMENSIONAL IMAGE
S511
SET FREQUENCY
S513
CUTOUT IMAGE
S515
SELECT MACHINE LEARNING MODEL
S517
GENERATE MACHINE LEARNING MODEL
S519
DISPLAY MACHINE LEARNING MODEL
S521
PROCESSING END?
NO
YES
END

FIG. 17

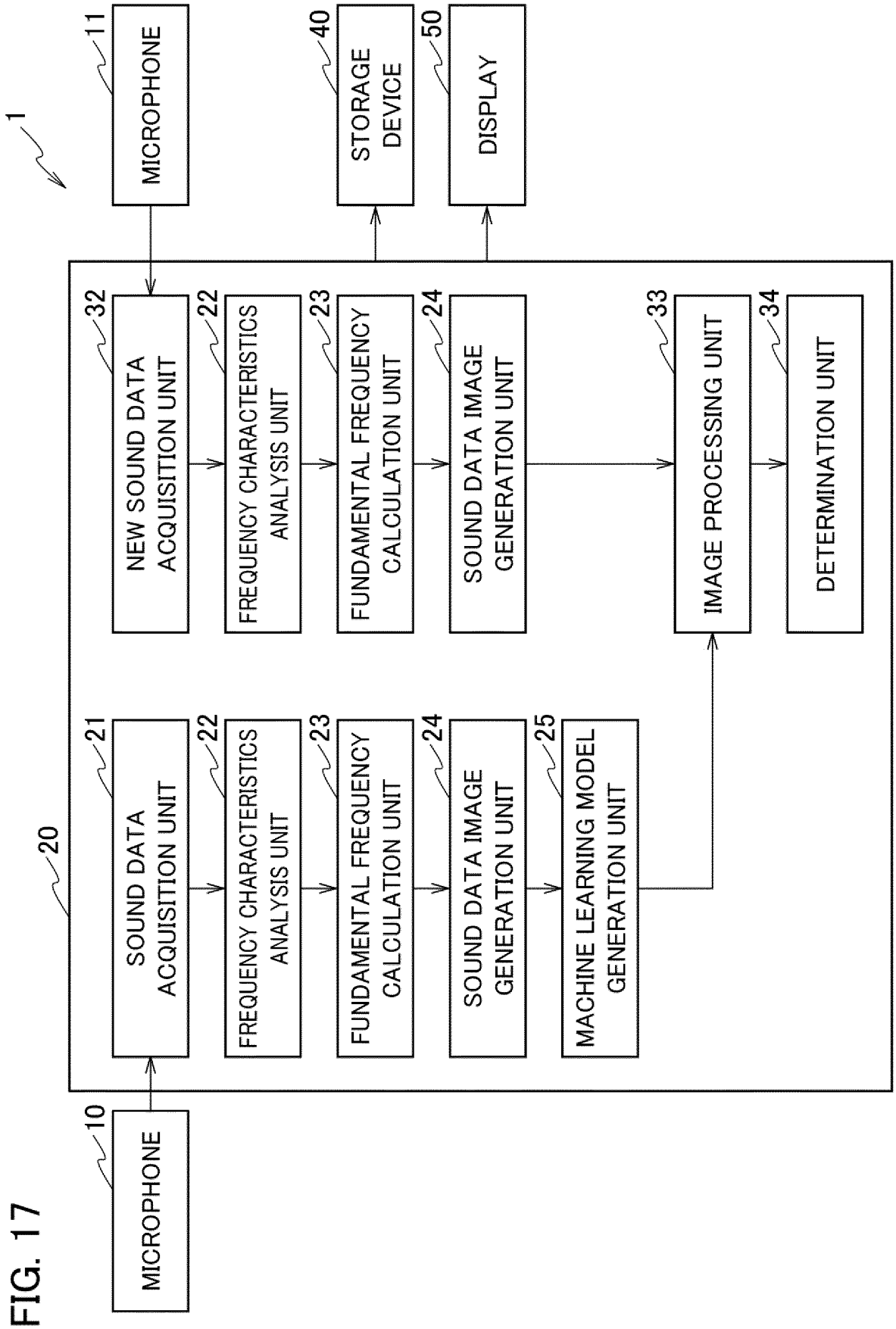

1
MICROPHONE
11
STORAGE DEVICE
40
DISPLAY
50
20
NEW SOUND DATA ACQUISITION UNIT
32
FREQUENCY CHARACTERISTICS ANALYSIS UNIT
22
FUNDAMENTAL FREQUENCY CALCULATION UNIT
23
SOUND DATA IMAGE GENERATION UNIT
24
IMAGE PROCESSING UNIT
33
DETERMINATION UNIT
34
SOUND DATA ACQUISITION UNIT
21
FREQUENCY CHARACTERISTICS ANALYSIS UNIT
22
FUNDAMENTAL FREQUENCY CALCULATION UNIT
23
SOUND DATA IMAGE GENERATION UNIT
24
MACHINE LEARNING MODEL GENERATION UNIT
25
MICROPHONE
10

FIG. 18A

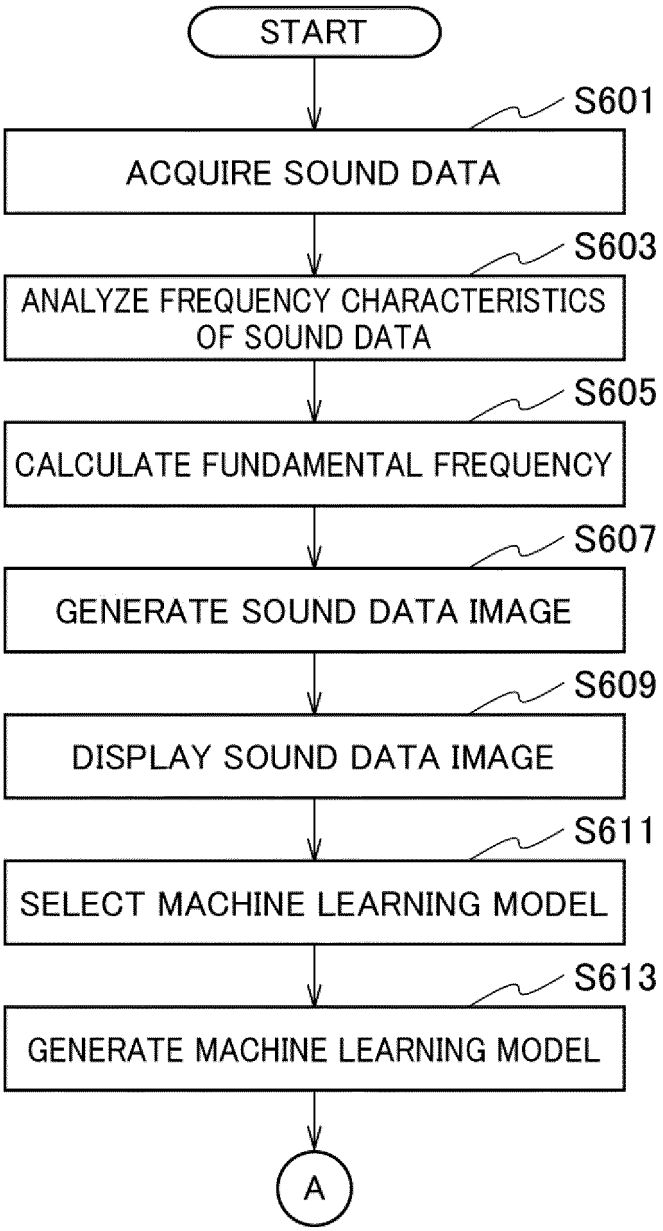
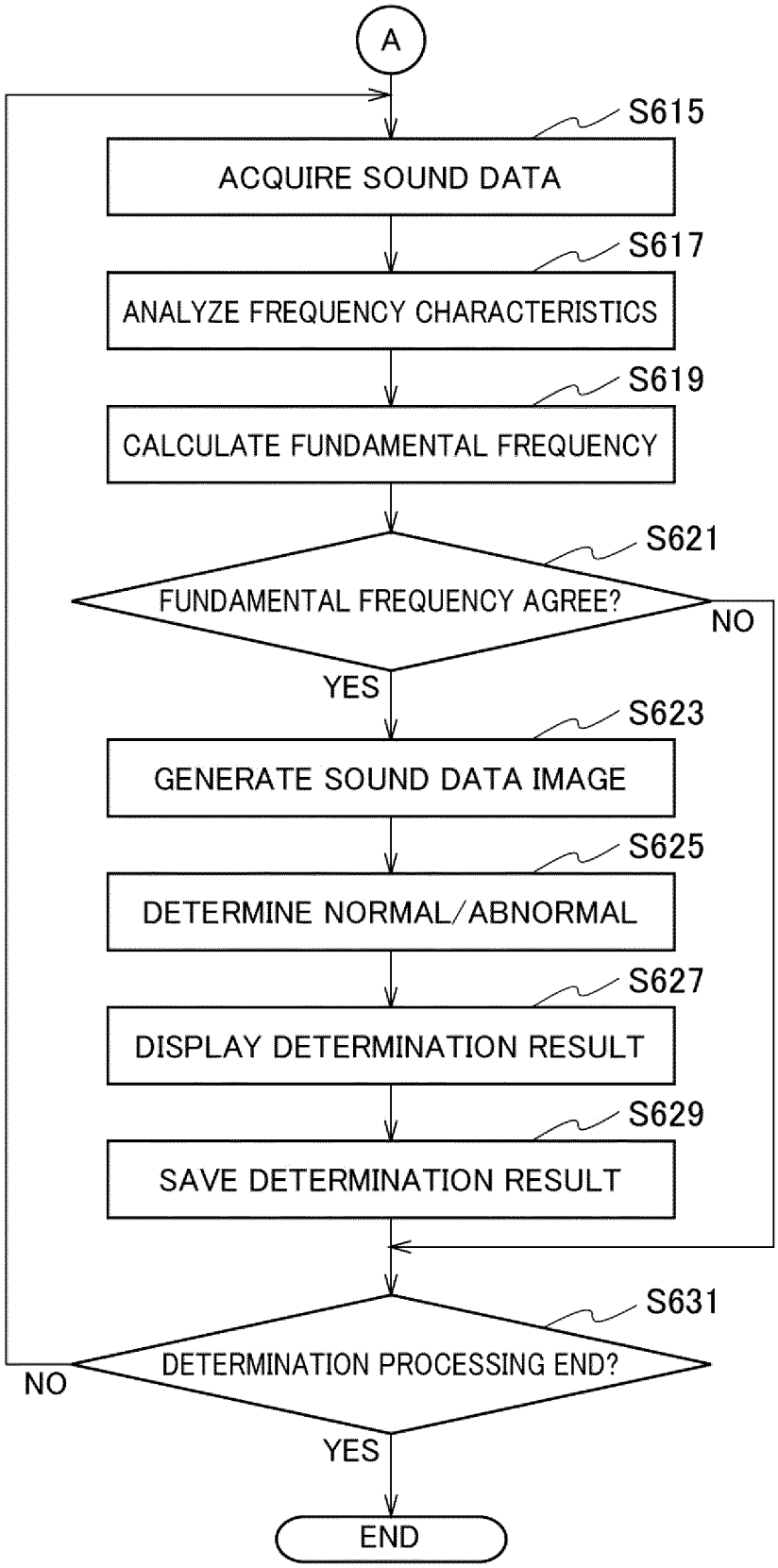

START
S601
ACQUIRE SOUND DATA
S603
ANALYZE FREQUENCY CHARACTERISTICS OF SOUND DATA
S605
CALCULATE FUNDAMENTAL FREQUENCY
S607
GENERATE SOUND DATA IMAGE
S609
DISPLAY SOUND DATA IMAGE
S611
SELECT MACHINE LEARNING MODEL
S613
GENERATE MACHINE LEARNING MODEL
A

FIG. 18B


A
S615
ACQUIRE SOUND DATA
S617
ANALYZE FREQUENCY CHARACTERISTICS
S619
CALCULATE FUNDAMENTAL FREQUENCY
S621
FUNDAMENTAL FREQUENCY AGREE?
NO
YES
S623
GENERATE SOUND DATA IMAGE
S625
DETERMINE NORMAL/ABNORMAL
S627
DISPLAY DETERMINATION RESULT
S629
SAVE DETERMINATION RESULT
S631
DETERMINATION PROCESSING END?
NO
YES
END

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IB2020/000097

A. CLASSIFICATION OF SUBJECT MATTER
Int. Cl. G01H3/08(2006.01)i, G01M99/00(2011.01)i, G06T7/00(2017.01)i, G10L25/30(2013.01)i, G01L25/51(2013.01)i
FI: G10L25/30, G01H3/08, G01M99/00 Z, G06T7/00 350B, G10L25/51 300

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
Int. Cl. G01H3/08, G01M99/00, G06T7/00, G10L25/30, G10L25/51

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Published examined utility model applications of Japan	1922-1996
Published unexamined utility model applications of Japan	1971-2020
Registered utility model specifications of Japan	1996-2020
Published registered utility model applications of Japan	1994-2020

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	WO 2019/176029 A1 (YAMAHA CORPORATION) 19 September 2019, paragraphs [0036], [0068]-[0069], [0071]	1-3, 5, 10 1-5, 7-8, 10
Y	JP 2013-076909 A (CASIO COMPUTER CO., LTD.) 25 April 2013, paragraphs [0010]-[0013], [0026]-[0032], [0044]-[0045]	1-5, 7-8, 10
Y	JP 2017-521705 A (FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E. V.) 03 August 2017, paragraph [0051]	4, 7-8

☒ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
09.06.2020	23.06.2020

Name and mailing address of the ISA/	Authorized officer
Japan Patent Office 3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan	Telephone No.

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IB2020/000097

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	坂本俊之．MXNet で作るデータ分析 AI プログラミング入門．初版．株式会社シーアンドアール研究所．02 July 2018, pp. 175-196, non-official translation (SAKAMOTO, Takayuki. Introduction to Programming Data Analysis AI Made with MXNet. First Edition. K.K. C&R Kenkyujo.)	8
A	JP 9-166483 A (HITACHI, LTD.) 24 June 1997, entire text, all drawings	1-10
A	木崎健太郎．振動監視に機械学習と振動発電，鉄道総研が不具合検出の実用研究．日経ものづくり．01 October 2018, no. 769, pp. 34-35, non-official translation (OZAKI, Kentaro. Machine Learning and Vibration Power Generation for Vibration Monitoring: Railway Technical Research Institute Launches Fault Detection Feasibility Study. Nikkei Monozukuri [Design].)	1-10

Form PCT/ISA/210 (continuation of second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IB2020/000097

Patent Documents referred to in the Report	Publication Date	Patent Family	Publication Date
WO 2019/176029 A1	19.09.2019	(Family: none)	
JP 2013-076909 A	25.04.2013	(Family: none)	
JP 2017-521705 A	03.08.2017	US 2017/0110135 A1 paragraph [0149] WO 2016/001067 A1 EP 2963645 A1 CN 106575510 A KR 10-201 7-0030549 A	
JP 9-166483 A	24.06.1997	(Family: none)	

Form PCT/ISA/210 (extra sheet) (January 2015)

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- WO 2015068446 A **[0003]**